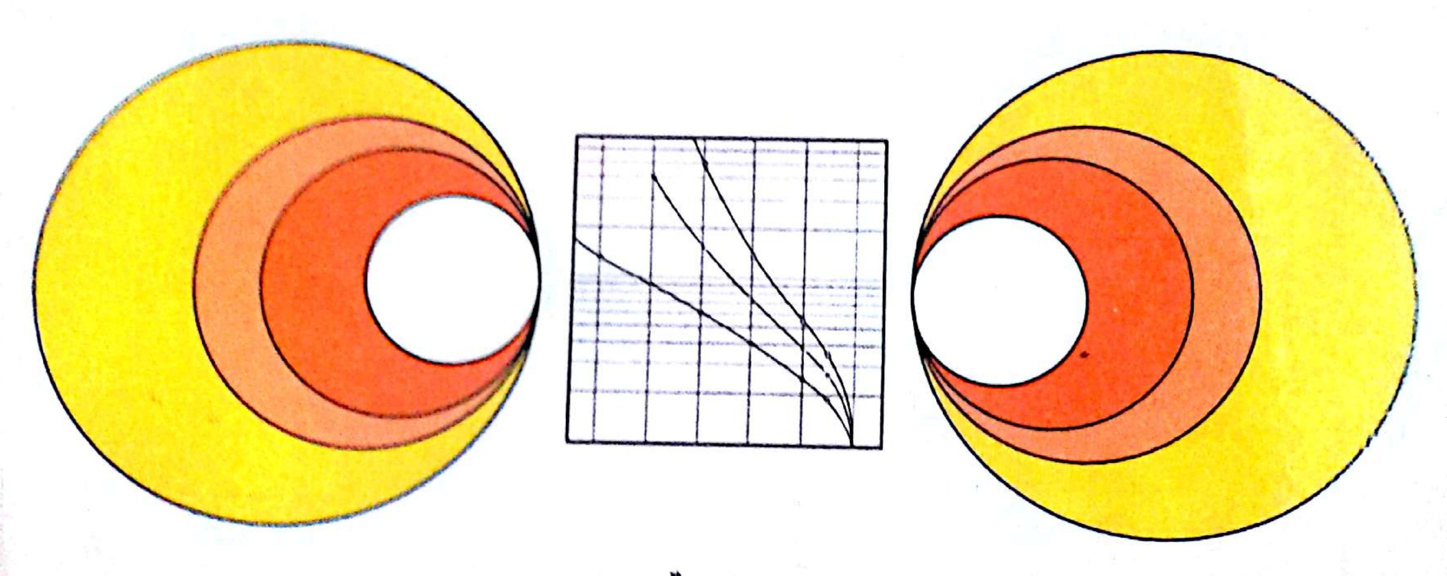
تقنية
المعاينة الإحصائية

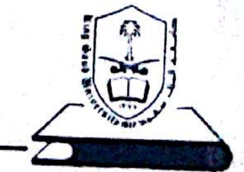
تاليف

ويليام كوكران



ترجمة

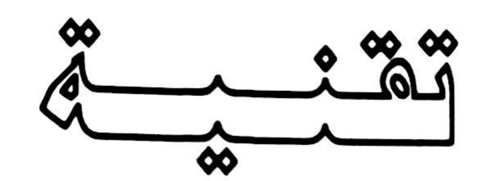
الدكتور أنيس كنجو



جامعة الملك سعود

عمادة شؤون المكتبات





المعاينة الإحصائية

الطبعة الثالثة

تأليف

ويليام كوكران

ترجمة

الدكتور أنيس كنجو

قسم الإحصاء وبحوث العمليات - كلية العلوم

جامعة الملك سعود



عمادة شؤون المكتبات - جامعة الملك سعود

ص.ب ٢٢٤٨٠ - الرياض ١١٤٩٥ - المملكة العربية السعودية

هذه ترجمة عربية مسموح بها لكتاب : Sampling Techniques, 3rd Edition

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

كوكـروان، ويليـام

تقنيـة المعاينـة الإحصائيـة/ ترجمة أنيـس كنجـو.

٦٣١ص، ١٧ × ٢٤ سم

ردمك ٥ ـ ٢٥٧ ـ ٠٥ ـ ٩٩٦٠ (غلاف)

٣ ـ ٢٥٨ ـ ٠٥ ـ ٩٩٦٠ (جلد)

١ ـ العيّنـات (إحصـاء) ١ ـ كنجـو، أنيـس (مترجم)

ب ـ العنوان

ديوي ٥١٩,٥٢ ١٦/٠٤٢٨

رقم الإيداع : ١٦/٠٤٢٨

حكمت هذا الكتاب لجنة متخصصة شكّلها المجلس العلمي بالجامعة، وقد وافق المجلس على نشره، بعد اطلاعه على تقارير المحكمين، في اجتماعه الثاني عشر للعام الدراسي ١٤٠٨/١٤٠٩هـ الـذي عُقـد بتـاريخ ١٤٠٩/٨/٥هـ الموافق ١٩٨٩/٣/١٢م.

مطابع جامعة الملك سعود ١٤١٦هـ

المحتويات

الصفحة

الفصل الخامس (ا): إضافات في أوجه المعاينة الطبقيّة

الفصل السادس: المقدِّر النسبة

الفصل الثامن : المعاينة النمطيّة

الفصل التاسع : المعاينة العنقوديّة وحيدة المرحلة : عناقيد متساوية الحجم

الفصل التاسع (أ): المعاينة العنقوديّة وحيدة المرحلة: عناقيد ذات حجوم غير متساوية

الفصل العاشر: المعاينة الجزئيّة بوحدات متساوية الحجم

الفصل الحادي عشر: المعاينة الجزئيّة بوحدات غير متساوية الحجم

* الفصل الثاني عشر: المعاينة المضاعفة

الفصل الثالث عشر: مصادر الخطأ في المسوح الإحصائية

مقدمة المترجم

الحمدلله وحده والصلاة والسلام على نبيِّنا محمد. وبعد، فلا يخفى ما لتطبيقات الإحصاء من دور متّسع ومهم في حياتنا المعاصرة. إذ يشكل الإحصاء اليوم إحدى أهم الأدوات المتوافرة للإنسان في سعيه الدؤوب للكشف عن المجهول في شروط تخضع للمصادفة. وفي ميدان التطبيقات الإحصائية تلعب المعاينة الإحصائية دورًا بارزًا. فهي تشكّل العمود الفقري لنشاطات المراكز الوطنية للإحصاء في كل بلد من بلدان العالم تقريبًا. ولها دور متميز في مراكز البحوث والدراسات حيثما وُجدت.

وإذا كانت المكتبة العربية تفتقر حتى إلى القليل من الكتب والمراجع في ميادين كثيرة من فروع العلوم المعاصرة، إلا أننا نكاد لانجد كتابًا مرجعيًّا واحدًا باللغة العربية في ميدان المعاينة الإحصائية. وإيمانًا منّا بضرورة حثّ الجهود لتعريب العلوم المعاصرة، وأن إضافة كتاب جديد إلى المكتبة العلمية العربية ينبغي له أن يتخذ شكل الواجب الحثيث تجاه ديننا وقومنا لكل مستطيع، فقد عزمت بعون من الله على ترجمة **«تقنية المعاينة»** لمؤلفه **ويليام كوكران**، ليكون إضافةً جديدة ميسّرة للدارس أو الباحث العربي. وعندما تخيّرت، بمشيئة الله، وقع اختياري على كتاب مشهور، لا يختلف اثنان في أنه أحد أفضل الكتب الموجودة في مجال المعاينة الإحصائية على مستوى العالم. ولمؤلفه باعٌ في تطوير نظرية المعاينة الإحصائية وتطبيقاتها الواسعة. وهو إذ يُشكّل مرجعًا قيّمًا في موضوعه، يشكل أيضًا كتابًا مدرسيًّا رائجًا تتبنّاه جامعات عديدة لطلابها

المتخصصين في العلوم الإحصائية سواءً في السنتين الأخيرتين من المرحلة الجامعية الأولى، أو في مرحلة الدراسات العليا.

وحرصًا على أن تخرج الترجمة في أفضل صورة فقد التمست من الأخ الأستاذ الدكتور عبدالرحمن أبوعمه مراجعة الترجمة. وتفضل مشكورًا بقراءة المخطوطة وزودني بالعديد من الملاحظات القيّمة خاصة فيما يتعلق بالمصطلح فجزاه الله كل الخير. وللزميلين اللذين قاما بتحكيم الترجمة جزيل الشكر على ما قدّماه من نصائح ومقترحات وملاحظات.

إن استكمال نواة أوليّة لمكتبة علمية يحتاج إلى جهود إضافية مضنية، وإلى أن يسود بين الاختصاصيين العلميين العرب شعور عميق بالمسؤولية والتقصير في آن واحد. فالزمن يمضي بسرعة ومكتبات العلوم في اللغات الحيّة، بمقياس اليوم، تزخر بزخم من الجديد في كل يوم وساعة. أما نحن الاختصاصيين العرب فنتوزّع بين اتكالي أناخ في بقيعة الاستسلام واليأس، لا يرى لنا مستقبلاً إلا من خلال الإنجليزية، أو الفرنسية، وبين متحمّس لرفد اللغة العربية، لغة الكتاب المنير، بكل ما يستطيع من حقائق العلوم المعاصرة وداعٍ إلى شدّ الهمم وتضافر الجهود، وبين لا مبالٍ أراح نفسه حتى من عناء التفكير في المشكلة. وكجزء من اهتمام واسع بتحقيق ذلك الحلم الكبير، حلم إرساء القواعد الأساسية لمكتبة علمية عربية، وحرصًا على الإسهام المتواضع في الجهود المبذولة على المستوى العربي للخروج بالطالب العربي من دائرة الحرمان والبؤس التي يعيشها وهو يبحث، دون جدوى، عن كتاب بلغته الأم يروي ظمأه إلى التزوّد بالعلم، ويخفّف من وطأة المعركة القاسية التي يخوضها لنيل المعرفة، أقدّم هذا الكتاب سائلاً الله العليّ القدير أن يتقبله مني عملاً صالحًا فهو من وراء القصد وهو الهادي إلى سواء السبيل.

المترجم

استهلال

يقدّم هذا الكتاب المدرسي، كما في الطبعات السابقة، وصفًا شاملاً لنظرية المعاينة في سياق تطوّرها الهادف إلى استخدامها في مسوح العيّنات. ويتضمن توضيحات تبيّن كيفية تطبيق النظرية في الميادين العلمية، وتمارين يحلُّها الطالب. وسيكون الكتاب مفيدًا ككتاب مدرسي لمقرّر في مسوح العيّنات يميل إلى التأكيد بصورة رئيسة على الجانب النظري، ولقراءات منفردة يقوم بها الطالب.

والحد الأدنى من التأهيل الرياضي الضروري لمتابعة الجزء الأعظم من مادة الكتاب هو معرفة ابتدائية بالجبر، وعلى وجه الخصوص، بعض العبارات الجبرية المعقدة نسبيًّا. بالإضافة إلى معرفة باحتمالات فضاءات العيّنة من النوع المنتهي، بما في ذلك احتمالات التوافيق. ويفترض الكتاب اجتياز الطالب لمقرر ابتدائي في الإحصاء يغطي المتوسطات والانحرافات المعيارية وبعض التوزيعات الاحتمالية مثل الطبيعي، والثنائي، وفوق الهندسي، ومتعدد الحدود، بالإضافة إلى نظرية النهاية المركزية، والانحدار الخطي، والأنواع المبسّطة من تحليل التباين. وبما أن نظرية المعاينة الكلاسيكية تتعامل في معظمها مع توزيعات المقدرات فوق مجموعة القيم العشوائية التي تقدمها خطة المعاينة، فستكون بعض المعرفة بطرق الإحصاء اللامعلمي مفيدة كذلك.

ومن حيث الأساس، قُدِّمت الموضوعات في هذه الطبعة بالترتيب نفسه الذي قُدِّمت فيه في الطبعات السابقة. وتتضمن هذه الطبعة موضوعات جديدة، أو فقرات أعيدت كتابتها، وذلك، بصورة رئيسة، لأحد أسباب ثلاثة:

(١) لتقديم مداخل إلى موضوعات (خطط معاينة أو طرائق تقدير) جديدة نسبيًا في حقل المعاينة.
(٢) لتغطية العمل الإضافي الذي تمّ خلال الخمس عشرة سنة الأخيرة على طرائق قديمة، واستهدف إما تحسين هذه الطرائق أو تعلُّم المزيد مما تؤديه طرائق بديلة.
(٣) تقصير أو توضيح أو تبسيط براهين أُعطيت في طبعات سابقة.

وتتضمن الموضوعات الجديدة في هذه الطبعة الطرائق التقريبية التي طُوِّرت لمعالجة مسألة صعبة هي مسألة حساب أخطاء معيارية أو وضع حدود ثقة لتقديرات غير خطية مأخوذة من نتائج مسوح إحصائية (مثلاً الانحدار)، وفي مسوح تتضمن أسئلة حسّاسة، لا يُحتمل أن تلقى الإجابة الصادقة عنها ترحيبًا من بعض المستجيبين، وهناك تدبير جديد، إذ نقدّم للمستجيب، بصورة عشوائية، السؤال الذي جرى تقديمه. وفي بعض مسائل المعاينة، يبدو استخدام قائمتين متداخلتين (أو ما يُسمى بالإطارين) لتغطية المجتمع بكاملة أمرًا مغريًا من الناحية الاقتصادية، أو أمرًا جوهريًا بالنسبة لبلدان لا تمتلك موارد معاينة كاملة. وقد عُمّمت طريقة المعاينة المضاعفة إلى حالات نستهدف فيها مقارنة متوسطات عدد من المجتمعات الجزئية ضمن المجتمع الأم. وهناك أعمال مفيدة تتعلق بالمزايا التي يتمتع بها التقدير النسبة وتقدير الانحدار إذا أمكن الافتراض بأن المجتمع المنته هو نفسه عيّنة عشوائية من مجتمع فوقي لانهائي يصحّ فيه نموذج رياضي مناسب للمقدّر النسبة أو لمقدّر الانحدار. ويُعتبر هذا النوع من الافتراض شيئًا جديدًا الآن (لاحظتُ حديثًا أن لابلاس قد استخدمه حوالي عام ١٨٠٠م في مسألة معاينة) إلا أنه يوضح العلاقة بين نظرية المعاينة ونظرية الإحصاء كما نعرفها اليوم.

وكمثال على بعض الأعمال الإضافية في موضوعات تضمنتها الطبعات السابقة نسوق الفصل ١٩ الذي كُتب في جزء منه من مادة كانت سابقًا في الفصل ٩؛ والهدف الرئيس من ذلك هو إعطاء وصف أكثر تلاؤمًا مع ما أعتقد أنه الطرائق الرئيسة التي استُخدمت لمعاينة غير متساوية الاحتمالات، وبدون إعادة. وهي تتضمن الطرائق

المتماثلة التي أُعطيت، بصورة مستقلة، من قبل Brewer ، و J.N.k. Rao ، Durbin ، وطـريقـة Murthy ، وطـريقـة Cochran, Hartley, Rao ، وطريقة Madow المتعلقة بالمعـاينـة النمطية، مع مقارنات لأداء هذه الطرائق في مجتمعات واقعية. وقد جرت دراسات جديدة لحجوم مركبات أخطاء القياس في المسوح الإحصائية وذلك بأسلوب تكرار القياسات بوساطة معايِنَين مختلفَين، وأسلوب العيّنات الجزئية المتداخلة، وعن طريق مركّب من الأسلوبين معًا. ومن أجل المقدر النسبة استُخدمت بيانات إحصائية من مجتمعات واقعية لتثمين تحيُّزات العيّنة الصغيرة في علاقات العيّنة الكبيرة المتعارف عليها والخاصة بالتباين، وبتقدير التباين. وقد بُذلت محاولات لابتكار أشكال جديدة أقل انحيازًا لمقدَّر النسبة نفسه وللعلاقة الخاصة بتقدير تباين المعاينة من أجله. وفي المعاينة الطبقية هناك عمل إضافي حول حصّة كل طبقة من الحجم الكلي للعيّنة وذلك عندما نهتم بأكثر من مفردة واحدة وحول تقدير أخطاء العيّنة عند اختيار وحدة واحدة فقط من كل طبقة. وقد نالت الاهتمام أيضًا بعض الطرائق الجديدة في المعاينة النمطية التي تتناول مجتمعات ذات نزعات خطية.

وقد أعدّ Alvx L. Fikner و Amil H. Jebe جزءًا كبيرًا من أمالي المحاضرات التي كُتبت منها الطبعة الأولى من هذا الكتاب. كما لقيت بعض الأبحاث، التي وفّرت مادّة خلفية، دعم مكتب البحث العلمي في سلاح البحرية ووزارة البحرية؛ ومن مناقشات حول التطورات الحديثة في المعاينة أو من اقتراحات حول هذه الطبعة، تلقيت عونًا كبيرًا من Leslie Kish, Daniel G. Horvitz, David J. Finney, Tore Dalinus, Amode R. Sen, Joseph Sedransk, Martin Sandelius, P.S.R. Sambasiva Rao . وبصورة خاصة John N.K. Rao ، الذي أدّت جهوده في قراءة الفقرات الجديدة والمحسَّنة لهذه الطبعة إلى العديد من الاقتراحات البنّاءة حول ثغرات ونقاط ضعف أو غموض، بالإضافة إلى اختيار المـوضـوعـات. وفيما يتعلق بالطباعة على الآلة الكاتبة والأعمال الأخرى التي ينطوي عليها إنتاج النسخة المعدّة للطبع فإني مدين لـ Rowena Foss ، Holly Grano و Edith Klotz شكري لهم جميعًا.

المؤلف

شباط (فبراير) ١٩٧٧



الفصل الأول

مقدمة

(١-١) فوائد طريقة العيّنة

تُبنى مواقفنا، معرفتنا، وأفعالنا، إلى حد كبير، على العيّنات. ويتساوى ذلك سواءً في الحياة اليومية العادية أو في البحث العلمي. وغالبًا ما يتحدد رأي شخص في مؤسسة تقوم يوميًا بآلاف الأعمال أو الإجراءات، من خلال ما واجهه لمرة أو مرتين، في هذه المؤسسة، خلال عدد من السنين. والمسافرون الذين يقضون 10 أيام في بلد أجنبي ثم يشرعون في إعداد كتاب يخبرون فيه سكان هذا البلد عن الكيفية التي يُنشّطون بها صناعتهم ويصلحون نظامهم السياسي، ويتلافون العجز في ميزانيتهم ويحسّنون الطعام في فنادقهم هم شخصيات فكاهية مألوفة. ويختلف هؤلاء، في حقيقة الأمر، عن الباحث في العلوم السياسية الذي يخصص 20 سنة للعيش والدراسة في ذلك البلد، فقط في أنهم يبنون نتائجهم على عيّنة من الخبرات أصغر بكثير، وفي كونهم أقل حظًّا في إدراك مدى جهلهم. وسواءً في العلوم أو في الأمور التي تخص البشر، تنقصنا الموارد التي تسمح لنا بدراسة ما يزيد على جزء صغير من الظواهر التي يمكن أن تدفع معرفتنا إلى الأمام.

ويحوي هذا الكتاب جردًا للجزء الرئيس من البحث النظري الذي تمّ بناؤه ليقدّم خلفية صالحة للوصول إلى طرق جيّدة للحصول على عيّنة. وفي معظم التطبيقات التي قام هذا البحث النظري من أجلها يكون المجتمع الإحصائي الذي نرغب في الحصول على معلومات حوله منتهيًا ومحددًا، سكان بلدة، الآلات في مصنع، السمك في بحيرة. . .

وقد يبدو معقولاً في بعض الحالات أن نحصل على معلومات دقيقة عن طريق تعداد تام أو حصر شامل للمجتمع. والإداريون الذين اعتادوا على التعامل مع عمليات الحصر الشامل كانوا، في البداية، يرتابون في العيّنات ويرفضون استخدامها بديلاً للحصر الشامل. ومع أن مثل هذا الموقف لم يعد قائمًا الآن، فقد يكون من المستحسن استعراض الفوائد الرئيسة لطريقة العيّنات، وذلك بالمقارنة مع التعداد التام.

أقل تكلفة

إذا توافرت المعلومات الإحصائية من جزء صغير فقط من المجتمع، فستكون النفقات أقل مما لو حاولنا القيام بتعداد تام. وفي مجتمعات كبيرة يمكن الحصول على نتائج، هي على قدر من الدقة يمكن معه اعتبارها مفيدة، وذلك من عيّنات تمثل فقط كسرًا صغيرًا من المجتمع. وفي الولايات المتحدة نجد أن عمليات المسح الإحصائي الأكثر تواترًا وأهمية والتي تقوم بها الحكومة تستخدم عيّنات تتضمن حوالي 105000 شخص، أي حوالي واحد من بين كل 1240 . ويمكن للمسوح الإحصائية المستخدمة لتقديم حقائق في أبحاث التسويق، تتعلق بالمبيعات وبسياسة الدعاية والإعلان، أن تستخدم عيّنات تتضمن آلافًا قليلة فقط.

أكثر سرعة

وللسبب نفسه يمكن جمع البيان الإحصائي وتلخيصه، عند أخذ عيّنة، بسرعة أكبر مما لو قمنا بتعداد تام. وهو أمر حيوي عندما تكون حاجتنا للمعلومات ملحّة.

أوسع أفقًا

في بعض أنواع الاستبيانات، لابد من استخدام أشخاص مدربين بصورة عالية أو جهاز مختص للحصول على المعلومات الإحصائية، وتوفّر مثل ذلك يكون محدودًا. وقد يكون الحصر الشامل، عندئذ، أمرًا غير عملي إذ أن الاختيار هو بين الحصول على المعلومات بوساطة العيّنات أو عدم الحصول على أية معلومات. وهكذا يكون للمسح

الإحصائي الذي يعتمد على العيّنة مجال أوسع للاستخدام ومرونة أكبر فيما يتعلق بأنواع المعلومات التي يمكن الحصول عليها. وعلى الوجه الآخر إذا أردنا معلومات دقيقة حول العديد من التقسيمات الفرعية أو القطاعات في المجتمع فإن حجم العيّنة الذي نحتاجه للقيام بالمهمة المطلوبة سيكون أحيانًا من الضخامة بحيث يُشكّل التعداد التام أفضل حل في هذه الحالة.

أكثر دقة

وبسبب إمكانية استخدام أشخاص ذوي كفاءة عالية، وامكانية إخضاعهم لتدريب مركّز، ولأن الإشراف الأدقّ على العمل الميداني ومعاملة البيانات تصبح، عند انخفاض حجم العمل، أمورًا ممكنة، فقد تقدّم العيّنة نتائج أكثر دقة من تلك التي يقدمها ذلك النوع من التعداد التام الذي يمكن القيام به عمليًا.

(١-٢) بعض استخدامات مسوح العيّنة

إن أكثر ما يصدم المراقب للتطورات الجارية في طرائق المعاينة فوق السنوات الخمس والعشرين الماضية هو التزايد السريع في عدد وأنواع المسوح المستخدمة في عمليات المعاينة. وينشر المكتب الإحصائي في الأمم المتحدة تقارير من وقت إلى آخر بعنوان sample surveys of current interest يتضمن مسوح عيّنة قامت بها الدول الأعضاء. ويتضمن تقرير 1968 مسوحًا من 46 دولة. ويستهدف العديد من هذه المسوح معلومات واضحة الأهمية بالنسبة للتخطيط القومي في موضوعات مثل الإنتاج الزراعي واستخدام الأرض، والبطالة وحجم القوّة العاملة، والإنتاج الصناعي، وأسعار البيع بالجملة والمفرّق، والحالة الصحيّة للشعب ومداخيل ومصروفات الأسرة. ولكن يمكن العثور أيضًا على استبيانات أكثر تخصصًا. فمثلاً: ترتيبات الأجازة السنوية (أستراليا)، أسباب الطلاق (المجر)، القروض والاستثمار في الريف (الهند)، استهلاك المنازل للماء (فلسطين المحتلة)، الاستماع إلى الراديو (ماليزيا)، قضاء العُطل (هولندا)، التركيب العُمري للبقر (تشيكوسلوفاكيا)، والوظائف الشاغرة (الولايات المتحدة).

وتلعب المعاينة دورًا بارزًا في الإحصاءات القومية العَشْرية وقد أُدخلت عيّنات الـ٥٪ في إحصاء 1940 في الولايات المتحدة وهي تتضمن أسئلة إضافية حول المهنة، النَّسبْ أو الأصل (parentage) ، وما شابه، وذلك بالنسبة للأشخاص الذين تقع أسماؤهم في سطرين من السطور الأربعين في كل صفحة من صفحات السجلّ. وقد اتّسع استخدام المعاينة كثيرًا في عام 1950 . فمن عيّنات الـ 20% (كل خامس سطر) تمَّ الحصول على معلومات تتعلق بأمور مثل الدخل، وسني المدرسة، والهجرة، والخدمة في القوات المسلحة، وبأخذ كل سادس شخص في عيّنات الـ 20% ، تشكلت عيّنة $3\frac{1}{3}$ % لتعطي معلومات حول الزواج والإنجاب. وقد قُسّمت سلسلة من الأسئلة المتعلقة بحالة وعمر شُقق السكن إلى خمس مجموعات وتُملأ الأجوبة في كل مجموعة لكل خامس منزل. وقد استخدمت المعاينة أيضًا للإسراع في نشر النتائج. وقد ظهرت الجداول التمهيدية للعديد من المفردات المهمة معدّة على أساس العيّنة، قبل سنة ونصف من ظهور التقارير النهائية.

وقد استمرت هذه الطريقة في تعدادي عام 1960 وعام 1970 . وباستثناء معلومات أساسية معيّنة مطلوبة من كل شخص لأسباب دستورية أو قانونية فقد تراجعت جميع عمليات التعداد الشامل لتصبح عمليات قائمة على أساس العيّنة.

وبالإضافة إلى استخدام العيّنات في عمليات الحصر الشامل تقوم المكاتب الحكومية باستخدامها بصورة مستمرة للحصول على آخر المعلومات. وكأمثلة من الولايات المتحدة نذكر "current population survey" الذي يقدم شهريًا بيانات حجم وتركيب القوة العاملة وحجم البطالة (أو العطالة)، كما نذكر المسح الإحصائي الصحي القومي، والسلسلة من العيّنات التي تمس الحاجة إليها لحساب الرقم القياسي الشهري لسعر المستهلك.

وعلى نطاق أضيق، تقوم الحكومات المحلية في مدينة أو ولاية أو منطقة باستخدام العيّنة، بصورة متزايدة، للحصول على معلومات تحتاجها من أجل التخطيط

للمستقبل، ولمواجهة مسائل ملحّة. وفي الولايات المتحدة نجد في معظم المدن الكبرى وكالات تجارية تقوم بأعمال التخطيط وتنفذ مسوح عيّنة لزبائنها.

وتعتمد بحوث التسويق كليًا على أسلوب المعاينة. وتقديرات حجوم مستمعي برامج مختلفة في الراديو والتليفزيون وقرّاء الصحف والمجلات (بما فيها الإعلانات الدعائية) تبقى جميعها تحت المراقبة الدقيقة والمستمرة. ويرغب رجال الصناعة وأصحاب محلات البيع بالمفرّق في معرفة مدى استجابة الناس لمنتجات جديدة أو طرق جديدة في الرّزم والتغليف (packaging)، ومعرفة شكاويهم من المنتجات القديمة، وأسباب تفضيلهم لسلعة على أخرى. وفي الصناعة وعالم الأعمال استخدامات عديدة لطرائق المعاينة في محاولة لزيادة كفاءة عملياتهم الداخلية. وتقع الميادين المهمة لضبط الجودة، وعيّنات القبول خارج نطاق هذا الكتاب. إلا أنه من الواضح أن القرارات المتخذة بالنسبة لمستوى الجودة أو لتغيّر في الجودة، أو لقبول أو رفض دفعات من البضائع لن يكون لها ما يبررها إلا إذا كانت النتائج التي حصلنا عليها من بيانات العيّنة صحيحة ومشروعة بالنسبة للدفعة بكاملها (ضمن حدود تساهل مقبولة). واستخدام طرائق المعاينة في سجلات العمليات التجارية (business transactions) (حسابات مالية، رواتب، المخزون من البضائع، شؤون الموظفين) ـ وأخذ عيّنة من السجلات يكون عادةً أسهل بكثير من أخذ عيّنة من الناس ـ يمكن أن يقدم، بسرعة وبتكلفة اقتصادية، معلومات مفيدة عمليًا. ويمكن تحقيق وفر باللجوء إلى المعاينة عند تقدير مخزون المستودعات، وفي دراسة لحالة جهاز معيّن وعمره، وعند التفتيش على دقة وإنتاجية عمّال المكاتب، وفي تقصّي الكيفية التي توزع بها شخصيّة رئيسة مهمة وقت عملها بين مهامها المختلفة، وبصورة أعم، في الحقل المسمى بحوث العمليات، وتتضمن كتب Deming (1960) و Solnim (1960) العديد من الأمثلة المفيدة التي تبيّن مدى تطبيقات طرائق المعاينة في حقل الأعمال. ودراسات سبر الرأي، والمواقف، والأصوات الانتخابية، التي قدّمت الكثير في مجال لفت نظر الجمهور إلى تقانة المعاينة، مازالت تشكل أمرًا له شعبيته في الصحف. وفي حقل الحسابات المالية وتدقيقها، الذي استخدم المعاينة لسنين عديدة، برز اهتمام جديد في مسألة تبنّي تطوّرات حديثة في مجال

المعاينة للمسائل الخاصة بهذا الحقل. وهكذا يصف Neter (1972)، كيف توفر شركات الخطوط الجوية والسكك الحديدية مالاً عند استخدامها لعيّنات من السجلات للتفريق بين الدخل الناجم عن الشحن والدخل الناجم عن نقل المسافرين. وقد خضع أيضًا وضع مسوح العيّنات كبيّنة في الدعاوي الحقوقية إلى مناقشة حيوية. وقد لاحظ Gallup (1972) الإسهام الرئيس الذي يمكن أن تقوم به طرائق المعاينة في عملية ترشيد الحكومة وذلك بتحديد آراء الناس بسرعة في برامج حكومية جديدة أو مقترحة، كما أكّد على دورها كمصادر معلومات في علم الاجتماع.

وبصورة عامة، يمكن تصنيف مسوح العينة إلى نوعين: وصفي وتحليلي. وفي مسح وصفي يكون الهدف ببساطة هو الحصول على معلومات معيّنة حول مجموعات ضخمة: مثلاً، عدد الرجال، والنساء، والأطفال الذين يشاهدون برنامجًا تلفازيًا معينًا. وفي مسح تحليلي، تكون المقارنة بين مجموعات جزئية مختلفة من المجتمع بغية اكتشاف ما إذا كانت هناك فروق بينها، ولصياغة، أو التحقق من، فرضيات تتعلق بأسباب هذه الفروق. والمسح المتعلق بالإنجاب في أنديانا بوليس، مثلاً، كان محاولة لتحديد مدى تخطيط الأزواج لعدد أطفالهم وللفترات الفاصلة بين ولادتين، وتحديد موقف الزوج والزوجة حيال هذا التخطيط، والأسباب الكامنة وراء تلك المواقف، ودرجة النجاح التي أحرزها مثل هذا التخطيط [Kiser and Whelpton, 1935].

والتمييز بين المسح الوصفي والمسح التحليلي ليس بالطبع فاصلاً. ويقدم العديد من المسوح بيانات تخدم كلي الهدفين. ومع تزايد عدد المسوح الوصفية برز على كل حال تزايد ملحوظ في المسوح التي قامت بصورة رئيسة لخدمة أهداف تحليلية، وبصورة خاصة في دراسة السلوك الإنساني والصحة، ويمكن أن نذكر هنا مسوحًا تتعلق بأسنان أطفال المدارس قبل وبعد إضافة الفلور إلى الماء، وبمعدّلات الوفاة وأسبابها بين الذين يدخنون بكميات مختلفة، والدراسات الضخمة المتعلقة بفعالية لقاح Salk ضد شلل الأطفال. ودراسة Coleman (1966) حول تكافؤ الفرص في مجال التربية، والتي نُفّذت على عيّنة من المدارس أخذت على مستوى قومي، واحتوت على العديد من تحليلات

الانحـدار التي قدرت الإسهـام النسبي لكـل من ميزات المدرسة، وخلفيّة المنزل، ومواصفات الطفل، في التغيّرات التي نجدها في نتائج الامتحان.

(١-٣) الخطوات الرئيسة في مسح عيّنة

تمهيدًا لمناقشة الدور الذي يلعبه الجانب النظري في مسح عيّنة، من المفيد أن نصف باختصـار الخطوات التي يتضمنها عادةً تخطيط مسح عيّنة إحصائي وتنفيذه. وتختلف المسوح الإحصائية، من حيث تعقيدها، اختلافًا كبيرًا. فأخذ عيّنة من 5000 بطاقة، مرقمة ومرتبة بصورة جيدة في ملف، هو عمل سهل. ولكنها مسألة أخرى أن تأخذ عيّنة من سكان منطقة يتم التنقل فيها بفضل أقنية مائية عبر الغابات، وحيث لا تتـوفـر خرائط ويتكلم السكان خمس عشرة لهجة مختلفة، وهم يرتابون في الغريب، ويرتابون جدًّا في غريب يوجه أسئلة. وقد تكون المشاكل التي تدعو للحيرة والارتباك في مسح إحصائي تافهة أو غير موجودة في مسح آخر.

ويمكن تجميع الخطوات الرئيسة في مسح إحصائي، على نحو كيفي إلى حدّ ما، تحت أحد عشر عنوانًا:

أهداف المسـح

يُشكِّل التعبير الواضح عن الأهداف أمرًا بالغ الأهمية. وبدون ذلك يمكن أن ننسى ـ عنـد الاستغراق في تفاصيل التخطيط ـ أهداف مسح معقد، وننتهي نتيجة لذلك إلى قرارات لا تتفق مع الأهداف.

المجتمع الذي سنأخذ منه العيّنة

وسنستخدم كلمة المجتمع للدلالة على مجمل المادة التي نختار منها العيّنة. وقد لا يقـدم تعـريف المجتمـع أي مشكلة عنـدمـا تكون العيّنة من دفعة من المصابيح الكهربائية المعدّة للإنارة بغية تقدير متوسط عمر المصباح. وعلى الوجه الآخر، عند معاينة مجتمع من المزارع لابد من وضع قواعد لتعريف مزرعة، وستبرز هنا دعاوى

تتعلق بالخط الذي يمثل حدود المزرعة. ويجب أن تكون هذه القواعد قابلة للاستخدام عمليًا: يجب أن يكون العدّاد قادرًا على أن يقرر في الحقل، وبدون الكثير من التردد، ما إذا كانت إحدى الحالات المريبة منتمية إلى المجتمع أم لا.

وحيثما يكون ممكنًا، فإنه ينبغي أن يتطابق المجتمع الخاضع للمعاينة مع المجتمع الذي نريد الحصول على معلومات عنه (المجتمع الهدف).

وأحيانًا، ولأسباب عملية أو توخيًا للسهولة، يقتصر المجتمع الخاضع للمعاينة على جزء من المجتمع الهدف، وإذا كان الأمر كذلك فينبغي أن نتذكر أن النتائج المستخلصة من العيّنة تنطبق على المجتمع الخاضع للمعاينة. والحكم على المدى الذي ستنطبق فيه هذه النتائج على مجتمع الهدف يجب أن يعتمد على مصادر معلومات أخرى. وأيّة معلومات إضافية يمكن جمعها حول طبيعة الفروق بين المجتمع الخاضع للمعاينة والمجتمع الهدف قد تكون مفيدة.

البيانات الإحصائية المراد جمعها

من الجيّد التحقق من أن جميع البيانات الإحصائية ملائمة للهدف من المسح الإحصائي وأنه لم تحذف أية بيانات أساسية. وغالبًا ما توجد نزعة لتوجيه الكثير من الأسئلة، وخاصةً في المجتمعات البشرية، وبعض هذه الأسئلة لا يجري تحليلها أبدًا فيما بعد. ويُخَفِّض الاستبيان الطويل جدًّا من دقة الأجوبة عن الأسئلة المهمة وغير المهمة على حدٍّ سواء.

درجة الدقة المطلوبة

تخضع نتائج مسوح العيّنة دائمًا لبعض الريبة، وهذا ناتج عن أن جزءًا فقط من المجتمع قد خضع للقياس، وبسبب أخطاء القياسات. ويمكن تخفيض هذه الريبة باللجوء إلى عيّنات كبيرة وباستخدام أجهزة قياس رفيعة المستوى. ويُكلِّف هذا، في العادة، وقتًا ومالاً. وبالتالي فإن تحديد درجة الدقة المطلوبة في النتائج هو خطوة مهمة. وهذه الخطوة هي مسؤولية الشخص الذي سيستخدم المعلومات الإحصائية. وقد يقدّم

تتعلق بالخط الذي يمثل حدود المزرعة. ويجب أن تكون هذه القواعد قابلة للاستخدام عمليًا: يجب أن يكون العدّاد قادرًا على أن يقرر في الحقل، وبدون الكثير من التردد، ما إذا كانت إحدى الحالات المريبة منتمية إلى المجتمع أم لا.

وحيثما يكون ممكنًا، فإنه ينبغي أن يتطابق المجتمع الخاضع للمعاينة مع المجتمع الذي نريد الحصول على معلومات عنه (المجتمع الهدف).

وأحيانًا، ولأسباب عملية أو توخيًا للسهولة، يقتصر المجتمع الخاضع للمعاينة على جزء من المجتمـع الهـدف، وإذا كان الأمر كذلك فينبغي أن نتذكر أن النتائج المستخلصة من العيّنة تنطبق على المجتمع الخاضع للمعاينة. والحكم على المدى الذي ستنـطبق فيه هذه النتـائـج على مجتمع الهدف يجب أن يعتمد على مصادر معلومات أخرى. وأيّة معلومات إضافية يمكن جمعها حول طبيعة الفروق بين المجتمع الخاضع للمعاينة والمجتمع الهدف قد تكون مفيدة.

البيانات الإحصائية المراد جمعها

من الجيّد التحقق من أن جميع البيانات الإحصائية ملائمة للهدف من المسح الإحصائي وأنه لم تحذف أية بيانات أساسية. وغالبًا ما توجد نزعة لتوجيه الكثير من الأسئلة، وخاصةً في المجتمعات البشرية، وبعض هذه الأسئلة لا يجري تحليلها أبدًا فيما بعـد. ويُخَفِّض الاستبيان الطويل جدًّا من دقة الأجوبة عن الأسئلة المهمة وغير المهمة على حدٍّ سواء.

درجة الدقة المطلوبة

تخضع نتائج مسوح العيّنة دائمًا لبعض الريبة، وهذا ناتج عن أن جزءًا فقط من المجتمع قد خضع للقياس، وبسبب أخطاء القياسات. ويمكن تخفيض هذه الريبة باللجوء إلى عيّنات كبيرة وباستخدام أجهزة قياس رفيعة المستوى. ويُكلِّف هذا، في العادة، وقتًا ومالاً. وبالتالي فإن تحديد درجة الدقة المطلوبة في النتائج هو خطوة مهمة. وهذه الخطوة هي مسؤولية الشخص الذي سيستخدم المعلومات الإحصائية. وقد يقدّم

مثل هذا التحديد بعض الصعوبات باعتبار أن العديد من الإداريين غير معتادين على صياغة أفكارهم بدلالة مقدار الخطأ في التقدير، الذي يمكن التسامح به، وبما لا يتنافى مع اتخاذ القرار الجيّد. وغالبًا ما يستطيع الإحصائي تقديم العون في هذه المرحلة.

طرائق القيـاس

قد تكون هناك اختيارات، سواء بالنسبة للأداة المستخدمة في القياس أو بالنسبة لطريقة الوصول إلى المجتمع. إذ يمكن، مثلاً، الحصول على معلومات حول الحالة الصحية لشخص من عبـارات يقـدمها الشخص نفسه، أو من فحص طبي. وقد يستخدم المسح استبيانًا يغني عن وجود شخص يوجه الأسئلة، أو معاينًا يقرأ مجموعة مألوفة من الأسئلة بدون حذر أو تمييز، أو عملية مقابلة تسمح بكثير من الحرية في شكل الأسئلة وترتيبها. ويمكن استخدام البريد أو الهاتف أو الزيارة الشخصية، أو مركّب من هذه الأساليب الثلاثة، وهناك الكثير من الدراسات حول طرائق ومشاكل المقابلة [انظر مثلاً: Hyman, 1954 و Payne, 1954].

ووضع بيانات تسجيل تكتب عليها الأسئلة وأجوبتها هو جزء رئيس من العمل التمهيدي. وفي الاستبيانـات البسيطة يمكن أحيانًا ترميز الأجوبة سلفًا أي كتابتها بطريقة يمكن معها تحويلها بصورة روتينية إلى التجهيزات الآلية. وفي الحقيقة، كي نضـع بيانـات تسجيل جيّدة، لابـدّ من تصـوّر بنية الجداول الملخصة النهائية التي ستُستخدم لاستخلاص النتائج.

الإطـار

قبل اختيار العيّنة يجب تقسيم المجتمع إلى أجزاء تسمى وحدات المعاينة، أو الـوحدات. ويجب أن تغطي وحدات المعاينة المجتمع بكامله، ولابد أن تكون هذه الوحدات منفصل بعضها عن بعض تمامًا، بمعنى أن كل عنصر من المجتمع ينتمي بالضبط إلى وحدة واحدة فقط. وأحيانًا تكون الوحدة المناسبة واضحة، كما هي الحال في مجتمع من المصابيح الكهربائية، حيث يشكل كل مصباح بمفرده وحدة. وهناك،

أحيانًا، اختيارات ممكنة بالنسبة للوحدة. فعند أخذ عيّنة من محصول زراعي، يمكن أن تكون الوحدة حقلاً، أو مزرعة، أو مساحة من الأرض نحدد شكلها وأبعادها كما نريد. وغالبًا ما تكون هذه القائمة من وحدات المعاينة، وتدعى إطارًا، إحدى المسائل العملية الرئيسة. ومن تجربة مُرّة، اكتسب رجال المعاينة موقفًا ناقدًا حيال قوائم تمّ جمعها بصورة روتينية لغاية ما. وقد وجد أن مثل هذه القوائم هي في الغالب غير كاملة، أو أنّها غير واضحة، أو أنّها تتضمن قدرًا مجهولاً من التكرار، بالرغم من التأكيدات بأن العكس هو الصحيح. وقد يكون من الصعب المجيء بإطار جيّد عندما يكون المجتمع متخصصًا، كما في مجتمعات وكلاء المراهنات على جياد السباق، أو الناس الذين يقومون بتربية دجاج الحبش. ويقدم Jessen, 1955 طريقة مفيدة لإقامة إطار من أغصان شجرة مثمرة.

اختيار العيّنة

يوجد الآن العديد من الخطط لاختيار عيّنة. ومن أجل كل من هذه الخطط يمكن القيام بتقدير أولي لحجم العيّنة، وذلك من معرفتنا بدرجة الدقة المرغوبة. وتجري أيضًا مقارنة التكاليف النسبية والزمن الذي تتطلبه كل خطة قبل اتخاذ قرار بتبنّي إحداها.

الاختبار المسبق

وقد وجد أنه من المفيد تجربة الاستبيان المقترح والطرق الميدانية على نطاق ضيق. إذ يُنتج هذا، على الدوام تقريبًا، تحسينات في الاستبيان، وقد يكشف عن مشاكل أخرى ستكون في نطاقها الواسع مشاكل جدّية، مثلاً أن نكتشف أن التكلفة ستكون أكبر بكثير مما توقعنا.

تنظيم عمل الميدان

نواجه في المسوح الإحصائية الواسعة العديد من مسائل إدارة الأعمال، إذ يجب أن يتلقى الأشخاص الذين سيعملون في المسح تدريبًا يتعلق بأهداف المسح وطرائق

القياس التي سيجري استخدامها، كما يجب أن يتوافر الإشراف المناسب عليهم أثناء عملهم. ووجود نظام للتدقيق المبكّر في نوعية العائدات الواردة في بداية المشروع أمر قيّم جدًّا. ويجب وضع الخطط لمعالجة حالات عدم الاستجابة، أي فشل العدّاد في الحصول على معلومات من بعض وحدات المعاينة.

تلخيص وتحليل البيان الإحصائي

الخطوة الأولى هي مراجعة الاستبيانات التي تمّ ملؤها بأمل تعديل الأخطاء الناتجة عن التسجيل أو ـ على الأقل ـ حذف المعلومات الإحصائية التي يتضح خطؤها. ونحتاج إلى اتخاذ قرارات تتعلق بطريقة الحساب في حالة إغفال المستجيب لبعض الأسئلة، أو حذف الأجوبة خلال عملية إعادة النظر المذكورة أعلاه. وبعد ذلك تتم الحسابات التي تقود إلى التقديرات المطلوبة. ومن أجل البيان الإحصائي نفسه تتوفر طرائق مختلفة في التقدير. ومن المستحسن عند تقديم النتائج الإفادة بمقدار الخطأ المتوقع في التقديرات الأكثر أهمية. وإحدى محاسن المعاينة الاحتمالية هي إمكانية وضع مثل هذه العبارات حول الخطأ المتوقع. مع أنه لابد من تقييدها بشدة إذا كان حجم عدم الاستجابة كبيرًا.

المعلومات المكتسبة لخدمة مسوح إحصائية مستقبلية

كلما كانت المعلومات التي تتوافر لنا عن المجتمع كثيرة منذ البداية، سهُل استنباط عيّنة تؤدي إلى تقدير دقيق. وإنجاز عيّنة يُشكّل، إلى حد كبير، مرشدًا لتحسين المعاينة مستقبلاً. وذلك من خلال المعلومات الإحصائية التي تقدمها حول المتوسطات، والانحرافات المعيارية، وطبيعة التغيّر في نتائج القياسات الرئيسة، وحول التكاليف المطلوبة للحصول على البيانات الإحصائية. وتتقدم عملية المعاينة بسرعة أكبر عندما نتخذ تدابير مسبقة لجمع معلومات من هذا النوع وتسجيلها.

وهناك ناحية مهمة أخرى يقدم فيها إنجاز عيّنة تسهيلات بالنسبة لعيّنات مستقبلية. ففي مسح إحصائي معقد لا تسير الأمور أبدًا وفق ما هو مخطط لها تمامًا.

والعدّاد الحذر يتعلم كيف يتعرف على الأخطاء في التنفيذ ويحرص على عدم وقوعها في مسوح إحصائية في المستقبل.

(١-٤) دور نظرية المعاينة

قدمنا هذه القائمة من الخطوات المطلوبة في مسح إحصائي للتأكيد على أن المعاينة عمل تطبيقي يستدعي عدة أنواع من المهارات المختلفة، وفي بعض الخطوات مثل تعريف المجتمع أو تحديد البيانات الإحصائية المراد جمعها وطرائق قياسها، أو تنظيم العمل الميداني تلعب نظرية المعاينة دورًا ثانويًا. ومع أن مثل هذه الموضوعات سوف لا تحظى بالمزيد من النقاش في هذا الكتاب إلا أنه ينبغي أن نأخذ أهميتها في الاعتبار. وتتطلب المعاينة الانتباه إلى جميع أطوار النشاط: فقد يحطم العمل الهزيل في طور ما مسحًا إحصائيًا أُنجز كل شيء آخر فيه بشكل جيّد.

وتهدف نظرية المعاينة إلى جعل المعاينة أكثر كفاءةً. وهي تحاول تطوير طرائق لاختيار العيّنة وطرائق للوصول إلى التقديرات التي ستنتجها العيّنة بأقل تكلفة ممكنة. هذه التقديرات التي نريدها دقيقة بما يكفي بالنسبة لهدفنا منها. وفي هذه الدراسة النظرية سنستخدم بصورة متكررة المبدأ الذي ينطلق من دقة محددة بأقل تكلفة ممكنة.

ولكي نُطبق هذا المبدأ، يجب أن نكون قادرين على التنبؤ بالدقة والتكلفة المتوقعة، وذلك في أي طريقة معاينة نأخذها بعين الاعتبار. وإلى الحد الذي يتعلق بالدقة، لا يمكننا، في أية حالة بعينها، التكهن بمقدار الخطأ المرتكب في تقدير ما، ذلك لأن مثل هذا الأمر يتطلب معرفة القيمة الحقيقية في المجتمع. وبدلاً من ذلك فإننا نحكم على دقة طريقة في المعاينة باللجوء إلى التوزيع التكراري لقيم التقدير التي سنحصل عليها لو أننا طبقنا الطريقة على المجتمع نفسه مرارًا وتكرارًا. وهذه هي بالطبع الطريقة المتّبعة عادةً في نظرية الإحصاء للحكم على الدقة.

وهناك تبسيط إضافي، ففي العيّنات ذات الحجم المستخدم عادةً في التطبيقات

العملية، تتوفر لنا، في الغالب، أسباب جيدة للفرض بأن التوزيع الاحتمالي للمقدّرات النـاتجة عن العيّنة هو، على وجه التقريب، التوزيع الطبيعي ومع مقدر يتوزع وفق التوزيع الطبيعي يكون شكل دالة التوزيع التكراري معروفًا بكامله إذا علمنا المتوسط والانحراف المعياري (أو التباين). ويهتم جزء كبير من نظرية المعاينة بإيجاد علاقات خاصة بهذه المتوسطات والتباينات.

وهناك فرقان بين النظرية القياسية للمعاينة (وهي النظرية المطبقة عمليًا)، وبين النظرية التقليدية للمعاينة كما تعلمها كتب الإحصاء الرياضي. ففي النظرية التقليدية نفـترض عادةً أن القياسات التي نقوم بها في وحدات المعاينة في المجتمع تتبع توزيعًا تكراريًا معينًا كالتوزيع الطبيعي، مثلاً، له شكل رياضي معروف، باستثناء معالم معيّنة للمجتمع مثل المتوسط والتباين لابد لنا من تقدير قيمها من بيانات العيّنة. أما في نظرية المعاينة، على الوجه الآخر، فيختلف الموقف إذ نفترض أن معلوماتنا عن توزيع التكرار محدودة للغاية. وبصورة خاصة، لا نفترض أن شكله الرياضي معروف، بحيث يمكن وصف مثل هذا الأسلوب بأنه حُر النموذج أو حر التوزيع. وهو الموقف الطبيعي بالنسبة لمسوح إحصائية ضخمة حيث نأخذ في كل وحدة معاينة العديد من القياسات ويكون لكل من هذه القياسات توزيع تكراري مختلف. وفي المسوح التي نقوم فيها بعدد قليل من القياسات فقط في كل وحدة، قد تبرر دراسة توزيعاتها التكرارية افتراض أشكال رياضية معروفة تسمح بتطبيق نتائج النظرية التقليدية.

والفرق الثاني هو أن المجتمعات في المسوح الإحصائية تتضمن عددًا منتهيًا من الوحدات. وتكون النتائج إلى حد ما أكثر تعقيدًا عندما تكون المعاينة من مجتمع منتهٍ بدلاً من مجتمع لانهائي. ومن الناحية العملية يمكن على الغالب تجاهل هذه الفروق في مجتمعات منتهية ولانهائية. وسنشير في حينه إلى الحالات التي لا يكون الأمر فيها كذلك.

(١-٥) المعاينة الاحتمالية

تمتلك جميع طرائق المعاينة، التي نستعرضها في هذا الكتاب الخواص الرياضية المشتركة التالية:

١ - نستطيع تعريف مجموعة العيّنات المتميزة $S_1, S_2, \ldots S_y$ التي يمكن لطريقة المعاينة أن تختارها عند تطبيقها على مجتمع معيّن، وهذا يعني أننا نستطيع أن نتعرف بدقة على وحدات العيّنة التي تنتمي إلى S_1 إلى S_2 وهكذا. وعلى سبيل المثال، لنفرض أن المجتمع يتألف من ست وحدات مرقمة من 1 إلى 6. فإحدى الطرق الشائعة لاختيار عيّنة حجمها 2 ترشح ثلاث عينات هي:

$$S_3 \sim (3, 6) \text{ ، } S_2 \sim (2, 5) \text{ ، } S_1 \sim (1, 4)$$

لاحظ أننا لا نحتاج إلى عرض جميع العيّنات الممكنة التي حجمها 2.

٢ - يُخصص لكل عيّنة ممكنة احتمال π_i واحتمال اختيارها من بين كل العيّنات الممكنة.

٣ - نختار واحدة من العيّنات S_i بطريقة عشوائية نخصص فيها لكل S_i الاحتمال المناسب لاختياره π_i. وفي المثال المذكور أعلاه يمكن أن نخصص احتمالات متساوية للعيّنات الثلاث. وعندئذ يمكن القيام بالسحب عن طريق اختيار عدد عشوائي بين 1 و 3. وإذا كان هذا العدد j نأخذ العيّنة S_j.

٤ - يجب تحديد طريقة لحساب التقدير من العيّنة، كما يجب أن تعود هذه الطريقة إلى تقدير وحيد من أي عيّنة محدّدة. وعلى سبيل المثال، يمكن أن نقرر أن التقدير هو متوسط القياسات التي حصلنا عليها من وحدات العيّنة.

ومن أجل أي طريقة للمعاينة تتصف بهذه الخواص، نكون في وضع يسمح لنا بحساب التوزيع التكراري للتقديرات التي تولّدها هذه الطريقة وذلك عند تطبيقها على المجتمع نفسه بصورة متكررة، ذلك لأننا نعلم التواتر الذي سنختار وفقًا له عيّنة محدّدة S_i. ومن الواضح إذن، أننا قادرون على تطوير نظرية معاينة لأي طريقة من هذا النوع، هذا بالرغم من أن تفاصيل هذا التطوير يمكن أن تكون معقدة، ويشير مصطلح «المعاينة الاحتمالية» إلى طريقة من هذا النوع.

وفي التطبيق، نادرًا ما نسحب عيّنة بكتابة العيّنات S_i والاحتمال π_i لكل منها، كما ذكرنا أعلاه. فهذا الأمر هو من المَشَقَّة بحيث يصبح مستعصيًا في المجتمعات الكبيرة التي يمكن أن تنتج بلايين من العيّنات الممكنة. وغالبًا ما يجري السحب بتحديد احتمالات الاختيار لكل من الوحدات ثم سحب هذه الوحدات، واحدة تلو الأخرى، أو في مجموعات، حتى تستكمل العيّنة بالحجم والنوع المرغوبين. وللغايات النظرية تكفي معرفة أننا نستطيع كتابة العيّنات S_i والاحتمالات π_i، إذا أردنا، بصرف النظر عن الوقت غير المحدود الذي نحتاجه لمثل ذلك.

(١-٦) بدائل المعاينة الاحتمالية

فيما يلي بعض الأنواع السائدة لمعاينة غير احتمالية.

١ - نقتصر عند أخذ العيّنة على جزء من المجتمع يمكن الوصول إليه بسهولة. فمثلًا يمكن أخذ عيّنة من الفحم من الجزء العلوي من عربة مقطورة تبلغ سماكته من 6 إلى 9 بوصات.

٢ - نختار العيّنة كيفما اتفق. عند التقاط 10 أرانب من قفص كبير في مختبر يمكن للباحث أن يأخذ تلك التي تقع يده عليها دون أي تخطيط واعٍ.

٣ - في حالة مجتمع صغير وغير متجانس يفتش المعاين كل المجتمع ويختار عيّنة صغيرة من الوحدات «النموذجية» - أي الوحدات القريبة مما يعتقد أنه يمثل المتوسط في المجتمع.

٤ - في دراسات تكون عملية القياس فيها غير سارة أو متعبة للشخص الخاضع للقياس، تتألف العيّنة، بصورة رئيسة من متطوعين.

وتحت شروط صحيحة يمكن لأي من هذه الطرق أن يعطي نتائج مفيدة ولكنها، على أي حال، لا تنقاد إلى معطيات نظرية المعاينة القائمة على أساس لا مَعْلمي، باعتبار أنه لا يوجد أي عنصر من العشوائية في عملية الاختيار. وربما كان الطريق الوحيد لامتحان مدى جودة كل من الطريقتين هو العثور على حالة تكون النتائج فيها معروفة، إما في المجتمع بكامله أو من خلال عيّنة احتمالية ثم القيام

بمقارنات. وحتى إذا بدا في إحدى هذه المقارنات تَفوّق طريقة ما فإن هذا لا يضمن تفوقها في ظروف أخرى مختلفة.

وفي هذا المضمار كانت بعض الاستخدامات المُبكّرة للمعاينة من قبل حكومات دول ومدن منذ عام 1850 وما بعد تهدف إلى تخفيض التكلفة باعتمادها لتقديرات عيّنة مأخوذة من عائدات تعداد إحصائي. ومن أجل المفردات الأكثر أهمية في التعداد كانت مجاميع الدولة أو المدينة تُحسب من بيانات التعداد الشامل. ولكن من أجل مفردات أخرى كان يجري اختيار عيّنة نسبتها 15 أو 25 بالمائة من عائدات التعداد كي تلقي الضوء على عملية تقدير مجاميع الدولة أو المدينة لهذه المفردات. وقد استخدمت طريقتان متنافستان لاختيار العيّنة. إحداهما وتدعى «اختيار عشوائي»، كانت تطبيقًا للمعاينة الاحتمالية التي يكون فيها لكل وحدة من وحدات المجتمع الفرصة نفسها في أن تؤخذ ضمن العيّنة. ومع هذه الطريقة تمّ التحقق أنه من خلال استخدام نظرية المعاينة والتوزيع الطبيعي، كما لاحظنا سابقًا، يمكن للمعاين أن يتنبأ من بيانات العيّنة، وبصورة تقريبية، بمقدار الخطأ المتوقع في التقديرات التي يستقيها من العيّنة. وفضلاً على ذلك يمكنه أن يتحقق، إلى حدما، من التنبؤات الخاصة بمعظم المفردات المهمة التي توافرت قيمها الواقعية من الحصر الشامل.

والطريقة الأخرى كانت «**الاختيار الهادف**». وهذه الطريقة لم تعرف على وجه التحديد بصورة مفصّلة إلا أن لها خاصتين بارزتين. فوحدة المعاينة تضمنت مجموعات من عائدات الحصر الشامل، وغالبًا ما تكون مجموعات ضخمة نسبيًا. وعلى سبيل المثال، في التعداد الإيطالي عام 1921 كانت البلاد تتألف من 8354 كومونة جرى تجميعها في 214 منطقة. وعند سحب عيّنة نسبتها 14% اختار الإحصائيان الإيطاليان Gini و Galvani 29 منطقة بصورة هادفة بدلاً من 1250 كومونة. ومن ثم فقد اختيرت المناطق التسع والعشرون بحيث تقدم العيّنة تقديرات دقيقة لسبعة متغيّرات مهمة كانت نتائجها معروفة على مستوى البلاد بأسرها. وكان الأمل أن هذه العيّنة قد تعطي أيضًا تقديرات جيّدة لمتغيّرات أخرى وثيقة الارتباط بهذه المتغيّرات السبعة.

وفي عام 1920 عيّن المعهد الدولي للإحصاء لجنة لتقدم تقريرًا عن محاسن ومساوىء الطريقتين. وقد بدا التقرير الذي قدمه Jensen (1926) بأنه على وشك تفضيل الاختيار الهادف. إلا أنه سرعان ما هُجرت طريقة الاختيار الهادف كأسلوب معاينة يهدف إلى الحصول على تقديرات قومية في مسوح إحصائية يكون عدد المفردات المقاسة فيها كبيرًا. إذ تنقصها المرونة التي أنتجتها تطوّرات لاحقة في المعاينة الاحتمالية، وكانت غير قادرة على أن تتنبأ من العيّنة بمقدار الدقة المتوقعة في التقديرات، كما أنها استخدمت وحدات عيّنة كبيرة جدًا. وقد انتهى Galvani و Gini إلى أن الطريقة الاحتمالية المسمّاة «**المعاينة العشوائية الطبقية**» (فصل ٥)، حيث وحدة المعاينة هي الكومونة كانت ستعطي نتائج أفضل من نتائج طريقتهما.

(١-٧) استخدام التوزيع الطبيعي

من المفيد أحيانًا استخدام كلمة مقدِّر للدلالة على قاعدة نحسب بموجبها، من نتائج العيّنة، تقديرًا لخاصة μ من خواص المجتمع، أما كلمة تقدير فتطلق على القيمة التي حصلنا عليها من عيّنة بالذات. وإذا كان $\hat{\mu}$ المقدِّر للمعلمة μ وفق خطة معاينة ما، فيسمّى $\hat{\mu}$ مقدارًا غير منحاز إذا كان متوسط قيمته المحسوبة من جميع العيّنات الممكنة للخطة مساويًا لقيمة المعلمة μ. ووفقًا لرموز الفقرة (١-٥) يمكن كتابة هذا الشرط على الشكل:

$$E(\hat{\mu}) = \sum_{i=1}^{v} \pi_i \hat{\mu}_i = \mu$$

حيث $\hat{\mu}_i$ هو التقدير الذي تعطيه العيّنة i والرمز E ويعني «قيمة توقع . . .» هو الرمز المستخدم عادةً.

وكما ذكرنا في الفقرة (١-٤) فإنّ العيّنات في المسوح الإحصائية هي، في الغالب، كبيرة بكفاية بحيث تخضع التقديرات الناتجة عنها تقريبًا للتوزيع الطبيعي. وأكثر من ذلك تتوافر لنا في حالة المعاينة الاحتمالية قوانين تعطينا متوسط التقديرات وتباينها. ولنفرض أننا أخذنا عيّنة بطريقة يُعرف عنها أنها تعطي تقديرًا غير منحاز وأننا حسبنا تقدير العيّنة $\hat{\mu}$ وانحرافه المعياري $\sigma_{\hat{\mu}}$ (غالبًا ما يُسمّى بصورة بديلة الخطأ المعياري)

فما هو مدى جودة التقدير؟ لا نستطيع أن نعرف القيمة المضبوطة لخطأ التقدير إلا أننا نعلم من خواص المنحنى الطبيعي أن الفرص هي :

0.32 (حوالي $\frac{1}{3}$) في أن تتجاوز القيمة المطلقة للخطأ $|\mu - \hat{\mu}|$ المقدار $\sigma_{\hat{\mu}}$.

0.05 (1 من 20) في أن تتجاوز القيمة المطلقة للخطأ $|\mu - \hat{\mu}|$ المقدار $1.96\sigma_{\hat{\mu}} = 2\sigma_{\hat{\mu}}$

0.01 (1 من 100) في أن تتجاوز القيمة المطلقة للخطأ $|\mu - \hat{\mu}|$ المقدار $2.58\sigma_{\hat{\mu}}$

وعلى سبيل المثال، إذا أعطت عيّنة احتمالية من سجلات بطاريات تُستخدم عادة، في مصنع كبير معدّل حياة $\hat{\mu}$ يساوي 394 يومًا مع خطأ معياري $\sigma_{\hat{\mu}}$ يساوي 4.6 يومًا فإن فرصة وقوع معدّل الحياة في مجتمع البطاريّات بين

$$\hat{\mu}_L = 394 - (2.58)(4.6) = 382 \text{ يومًا}$$

و

$$\hat{\mu}_U = 394 + (2.58)(4.6) = 406 \text{ يومًا}$$

هي 99 في المائة.

ويدعى الحدان 382 يومًا و 406 يومًا حدي الثقة الأدنى والأعلى. ومن أجل تقدير بمفرده من عيّنة بمفردها لا تكون صحة العبارة «μ يقع بين 382 و 406 يومًا» مؤكدة. والرقم « 99% ثقة» يتضمن أنه إذا استخدمنا في مجتمع ما، خطة المعاينة نفسها مرات عديدة ووضعنا عبارة ثقة من كل عيّنة فقد يكون 99% من هذه العبارات صحيحًا و 1% خطأ. وعندما نستخدم المعاينة في عملية تمّ من أجلها تعدادات شاملة في الماضي يجري أحيانًا تحقيق هذه الخاصة بسحب عيّنات متكررة، من النوع المقترح، من مجتمع تتوفر له سجلات متكاملة، أي تكون قيمة μ معروفة [انظر مثلاً Cyert, 1957 Trueblood and]. ومثل هذا التحقق العملي من أن النسبة المعروضة من العبارات هي نسبة صحيحة، على وجه التقريب، يقدم الكثير في مجال تثقيف وتطمين الإداريين حول طبيعة المعاينة. وبصورة مماثلة، عند أخذ عيّنة بمفردها من كل من سلسلة من المجتمعات المختلفة فإن 95% من عبارات الـ 95% ثقة تكون صحيحة.

وتفترض المناقشة السابقة أن $\sigma_{\hat{\mu}}$ المحسوب من العيّنة معروف بالضبط وفي الواقع فإنّ $\sigma_{\hat{\mu}}$ مثله مثل $\hat{\mu}$ يخضع لخطأ المعاينة. وفي حال متغيّر يتوزع وفق التوزيع

الطبيعي، تستخدم جداول توزيع ستيودنت أو التوزيع t بدلاً من جداول التوزيع الطبيعي لحساب حدّي الثقة لـ μ، وذلك عندما تكون العيّنة صغيرة. واستبدال الجدول t بالجدول الطبيعي لا يحدث أي فرق تقريبًا إذا تجاوز عدد درجات الحرية في $\sigma_{\hat{\mu}}$ الـ 50. ومع أنواع معيّنة من المعاينة الطبقية، كما في طريقة المعاينة المكررة (فقرة ١١ - ١٩)، يكون عدد درجات الحرية صغيرًا ونحتاج إلى الجدول t.

(١ - ٨) الانحياز وتأثيره

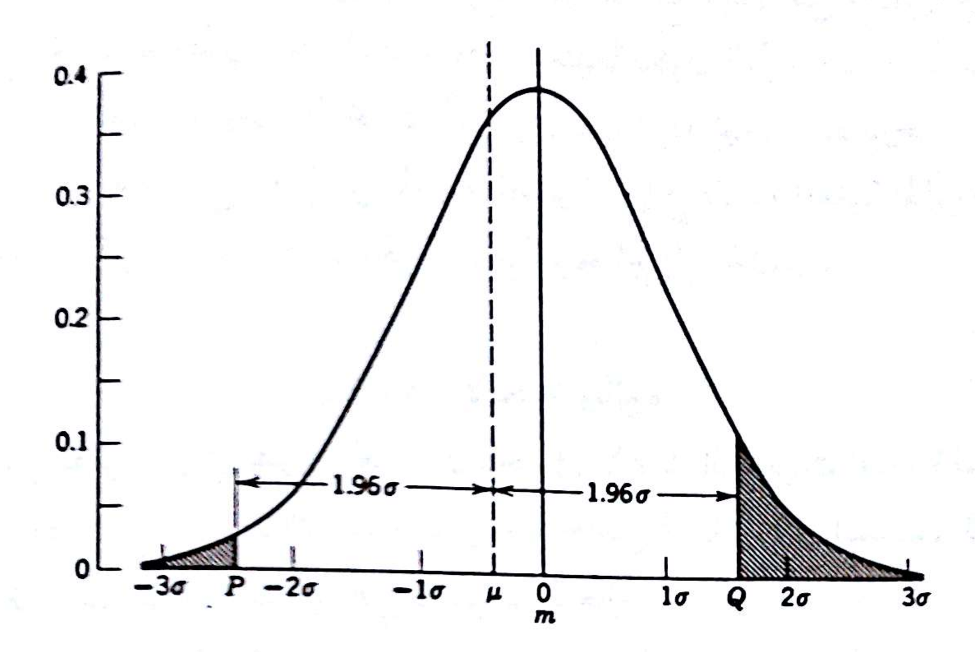
من الضروري في نظرية المعاينة أن نأخذ في الاعتبار التقديرات المنحازة لسببين:

١ - في بعض المسائل الأكثر شيوعًا، وبوجه خاص في تقديرات النسب، نجد أن التقديرات المريحة والمناسبة هي تقديرات منحازة.

٢ - وفي حالة المعاينة الاحتمالية نجد أنه حتى مع التقديرات غير المنحازة يمكن أن تُنتج أخطاء القياس وعدم الاستجابة انحيازًا في الأعداد التي نستطيع حسابها من بيانات العيّنة. ويحصل هذا، على سبيل المثال، إذا كان الأشخاص الذين رفضوا إجراء مقابلة كلهم تقريبًا من المعارضين لنوع صرف الأموال العامة، بينما ينقسم أولئك الذين أجروا المقابلة بالتساوي بين مؤيد ومعارض.

ولدراسة تأثير الانحياز، لنفرض أن التقدير $\hat{\mu}$ يتوزع طبيعيًا حول متوسط m يقع على مسافة B من المتوسط الحقيقي للمجتمع μ، وذلك كما هو مبين في الشكل (١-١). فمقدار الانحياز هو $B = m - \mu$. لنفرض أننا لا نعلم بوجود أي انحياز. ولنحسب الانحراف المعياري σ للتوزيع التكراري الموافق للتقدير، وسيكون هذا بالطبع الانحراف المعياري حول المتوسط المفترض للتوزيع m وليس حول المتوسط الحقيقي μ، نستخدم هنا σ بدلاً من $\sigma_{\hat{\mu}}$ وكعبارة حول دقة التقدير نقول إن احتمال أن يتجاوز خطأ التقدير $\hat{\mu}$ الكمية $1.96\,\sigma$ هو 0.05 فقط.

وسنرى كيف يحرف وجود الانحياز هذا الاحتمال. ولتوضيح ذلك نحسب الاحتمال الصحيح لكون الخطأ المرتكب في التقدير أكبر من $1.96\,\sigma$ حيث نقيس الخطأ



شكل (١-١) تأثير الانحياز على أخطاء عملية التقدير.

اعتبارًا من المتوسط الحقيقي μ . ولابد من تأمل كل من ذيلي التوزيع على حدة. فبالنسبة للذيل الأعلى، يكون احتمال وجود خطأ بالزيادة أكبر من $1.96\,\sigma$ وهو المساحة المظللة على يمين Q في الشكل (١-١). وهذه المساحة تساوي:

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\mu+1.96\sigma}^{\infty} e^{-(\hat{\mu}-m)^2/2\sigma^2}\, d\hat{\mu}$$

لنضع $\hat{\mu}-m=\sigma t$. فعندئذ يصبح الحد الأدنى للتكامل بدلالة t هو:

$$\frac{\mu-m}{\sigma}+1.96=1.96-\frac{B}{\sigma}$$

وتصبح المساحة على الشكل:

$$\frac{1}{\sqrt{2\pi}}\int_{1.96-(B/\sigma)}^{\infty} e^{-t^2/2}\, dt$$

وبصورة مماثلة فإن الذيل الأدنى، أي المساحة المظللة على يسار P ، تساوي:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-1.96-(B/\sigma)} e^{-t^2/2}\, dt$$

ويتضح من شكل التكاملين أن مقدار الاضطراب يعتمد فقط على نسبة الانحياز إلى الانحراف المعياري . والنتائج مبيّنة في الجدول (١-١) .

جدول (١-١) تأثير انحياز مقداره B على احتمال وجود خطأ أكبر من $1.96\ \sigma$.

B/σ	احتمال وجود خطأ $< -1.96\sigma$		المجموع
	$< -1.96\sigma$	$> 1.96\sigma$	
0.02	0.0238	0.0262	0.0500
0.04	0.0228	0.0274	0.0502
0.06	0.0217	0.0287	0.0504
0.08	0.0207	0.0301	0.0508
0.10	0.0197	0.0314	0.0511
0.20	0.0154	0.0392	0.0546
0.40	0.0091	0.0594	0.0685
0.60	0.0052	0.0869	0.0921
0.80	0.0029	0.1230	0.1259
1.00	0.0015	0.1685	0.1700
1.50	0.0003	0.3228	0.3231

وفيما يتعلق بالاحتمال الكلي لوجود خطأ أكبر من $1.96\ \sigma$ ، نجد أن للانحياز تأثيراً طفيفًا شريطة أن يكون هذا الانحياز أقل من عشر الانحراف المعياري . وعندما يكون الانحياز مساويًا لِـ $0.10\ \sigma$. فإن الاحتمال الكلي يكون 0.0511 بدلاً من 0.05 كما نظن . وكلما ازداد الانحياز يصبح الاضطراب أكثر خطورة . وعندما يكون $B = \sigma$ فإن الاحتمال الكلي يصبح 0.17 أي أكثر من ثلاثة أمثال القيمة التي نظن .

ويختلف الذيلان في تأثرهما . فمع الانحياز الموجب، كما في المثال المبين أعلاه، ينكمش احتمال تقدير بالنقصان يتجاوز 1.96σ ، انكماشًا سريعًا عن القيمة المفترضة 0.025 ، ليصبح مهملاً عندما $B = \sigma$. واحتمال التقدير بالزيادة المقابل يصعد بثبات . ويكون للخطأ الكلي الأهمية الأولى في معظم التطبيقات . ولكننا، من حين لآخر، نهتم بصورة خاصة بالأخطاء في اتجاه معين .

وكقاعدة عمل، نقول إن تأثير الانحياز على دقة تقدير ما يكون مهملاً إذا كان الانحياز أقل من عشر الانحراف المعياري لهذا التقدير. وإذا كانت لدينا طريقة منحازة في التقدير وكان $B/\sigma < 0.1$ حيث B القيمة المطلقة للتحيز، فيمكننا الادّعاء عندئذ بأن الانحياز لا يشكل عيبًا يُذكر لهذه الطريقة. وحتى في حالة $B/\sigma = 0.2$. يكون الاضطراب في احتمال الخطأ الكلي متواضعًا.

وعند استخدام هذه النتائج لابد من التمييز بين مصدري الانحياز المذكورين في بداية هذه الفقرة. ومع انحيازات من النوع الذي يبرز في تقدير النسب، يمكن أن نحسب بصورة رياضية حدًّا أعلى للنسبة B/σ. وإذا كانت العيّنة كبيرة بقدر كافٍ يمكن الاطمئنان إلى أن B/σ سوف لا تتجاوز 0.1. ومع انحيازات ناتجة عن أخطاء في القياسات أو عدم استجابة، يكون من المستحيل، على الوجه الآخر، إيجاد حد أعلى مضمون وصغير لـ B/σ. ونناقش هذه المسألة الشائكة في الفصل الثالث عشر.

(١-٩) متوسط مربعات الخطأ

كي نقارن تقديرًا غير منحاز، أو تقديرين منحازين بمقدارين مختلفين من الانحياز، يمكن استخدام قاعدة مفيدة هي متوسط مربعات خطأ التقدير (MSE)، مقاسًا بدءًا من قيمة المجتمع التي نريد تقديرها فنكتب:

$$\begin{aligned}
\text{MSE}(\hat{\mu}) = E(\hat{\mu}-\mu)^2 &= E[(\hat{\mu}-m)+(m-\mu)]^2 \\
&= E(\hat{\mu}-m)^2 + 2(m-\mu)E(\hat{\mu}-m) + (m-\mu)^2 \\
&= (\text{variance of } \hat{\mu}) + (\text{bias})^2
\end{aligned}$$

= (تباين $\hat{\mu}$) + ^{2}(الانحياز)

وينعدم الحد الجدائي باعتبار أن $E(\hat{\mu}-m)=0$.

واستخدام الـ MSE كقاعدة لدقة مقدَّر يؤدي إلى اعتبار تقديرين لهما الـ MSE نفسه كتقديرين متكافئين. وهذا ليس صحيحًا تمامًا باعتبار أن التوزيعين التكراريين لخطأين $(\hat{\mu}-\mu)$ مختلفين في حجميهما سوف لا يتطابقان من أجل التقديرين

إذا اختلفا في مقدار انحيازهما. إلا أن Hansen; Hurwitz & Madow (1953)، بيّنوا أنه إذا كان B/σ أقل من حوالي النصف فإن توزيعيْ التكرار يتطابقان تقريبًا فيما يتعلق بالقيم المطلقة لخطأين $|\hat{\mu}-\mu|$ من حجمين مختلفين. ويوضح الجدول ١-٢ هذه النتيجة.

جدول (١-٢) احتمال خطأ مطلق أكبر أو يساوي $1\sqrt{MSE}$, $1.96\sqrt{MSE}$ and $2.576\sqrt{MSE}$

الاحتمال

B/σ	$1\sqrt{MSE}$	$1.96\sqrt{MSE}$	$2.576\sqrt{MSE}$
0	0.317	0.0500	0.0100
0.2	0.317	0.0499	0.0100
0.4	0.319	0.0495	0.0095
0.6	0.324	0.0479	0.0083

وحتى إذا كان $B/\sigma=0.6$، فإن التغيرات في الاحتمالات بالمقارنة مع الاحتمال الموافق للحالة $B/\sigma=0$ هي تغيّرات طفيفة.

وبسبب صعوبة التأكد من عدم وجود انحياز أكيد، في التقديرات، سنتكلم عادة عن إحكام تقدير بدلاً من دقة تقدير. فالدقة تشير إلى حجم الانحراف عن المتوسط الصحيح μ، بينما يشير الإحكام إلى حجم الانحراف عن المتوسط m الناتج عن تطبيق أسلوب المعاينة نفسه بصورة متكررة.

تمارين

(١-١) لنفرض أنك تستخدم المعاينة لتقدير العدد الإجمالي للكلمات في كتاب يحوي توضيحات.

(أ) هل توجد أية مشكلة في تعريف المجتمع؟

(ب) ما هي الحجج المؤيدة والمضادة لاعتبار: (١) الصفحة، (٢) السطر، كوحدة معاينة؟

(١-٢) سنأخذ عيّنة من قائمة من الأسماء المسجلة على بطاقات (اسم في كل بطاقة) مرقمة على التسلسل في إضبارة. ولكل اسم الفرصة نفسها في أن يُسحب في العيّنة. ما هي المشكلات التي تنشأ في الحالات العامة التالية؟

(ا) بعض الأسماء لا تنتمي إلى المجتمع الهدف، علمًا بأن هذه الحقيقة لا يمكن التأكد منها إلا بعد سحب الاسم.

(ب) بعض الأسماء تظهر على أكثر من بطاقة، وجميع البطاقات التي تحمل الاسم نفسه لها أرقام متسلسلة وبالتالي تظهر بعضها إلى جانب بعض في الإضبارة.

(جـ) بعض الأسماء تظهر على أكثر من بطاقة والبطاقات التي تحمل الاسم نفسه مبعثرة ضمن الإضبارة.

(١-٣) تشكل مسألة إيجاد إطار كامل يسمح بسحب العيّنة عقبة في الغالب. ما هي أنواع الإطارات التي يمكن تجربتها في المسوح التالية؟ وهل لهذه الإطارات أية نقاط ضعف خطيرة؟

(ا) مسح للمخازن التي تبيع حقائب سفر في مدينة كبيرة.

(ب) مسح لأنواع المواد التي تركها أصحابها في قطارات الأنفاق أو الحافلات العامة.

(جـ) مسح للأشخاص الذي لدغتهم الثعابين في العام الماضي.

(د) مسح لتقدير عدد الساعات الأسبوعية التي يقضيها أفراد أسرة في مشاهدة التلفاز.

(١-٤) دليل مدينة عمره أربع سنوات ويتضمن العناوين مرتّبة على طول كل شارع، كما يعطي أسماء الأشخاص الذين يعيشون في كل عنوان. يُراد إجراء مسح للأشخاص في المدينة، يجري بطريقة المقابلة، ما هي نواقص هذا الإطار؟ هل يمكن معالجة هذه النواقص من قبل العدّادين خلال قيامهم بعملهم الميداني؟ عند استخدامك للدليل، هل تسحب قائمة من العناوين (أمكنة السكن) أم قائمة من الأشخاص؟

(١-٥) عند تقدير القيمة الفعلية للبنود الصغيرة في مستودعات شركة كبيرة

بطريقة العينة، سجّلنا القيمة الفعلية إلى القيمة الدفترية لكل البنود في العيّنة. ووجدنا أن نسبة القيمة الفعلية إلى القيمة المسجّلة في العيّنة كلها كانت 1.021 ، وهذا التقدير يتوزع طبيعيًا بانحراف معياري 0.0082 . إذا كانت القيمة الدفترية للمخزون المراد تقدير قيمته الفعلية هي 80,000 $ ، فاحسب 95% حدود ثقة للقيمة الفعلية.

(١-٦) كثيرًا ما نتعامل مع بيان إحصائي كعيّنة، مع أنها تبدو للوهلة الأولى وكأنها حصر شامل. ويجد صاحب موقف للسيارات ضآلة العمل في أيام الأحد صباحًا. وبعد ستة وعشرين يوم «أحد» من العمل كان متوسط ما تسلمه صباح الأحد هو 10 $ بالضبط. والخطأ المعياري لهذا الرقم، محسوبًا من التغيّرات بين أسبوع وآخر هو 1.2 $. ويتقاضى الحارس 7 $ كل أحد. ويُرحِّب المالك بترك الموقف مفتوحًا للسيارات صباح الأحد إذا كان توقع ربحه في المستقبل يبلغ الـ 5 $ كل صباح أحد. ما معامل الثقة الاحتمالية بأن معدّل ربحه على المدى الطويل سيكون 5 $ على الأقل؟ ما الفرضيات التي يجب وضعها للإجابة عن هذا السؤال؟

(١-٧) في الجدول (١-٢) ماذا يحدث لاحتمالات تجاوز $1\sqrt{MSE}$ ، $1.96\sqrt{MSE}$ و$2.576\sqrt{MSE}$ ، عندما ينتهي B/σ إلى اللانهاية، أي عندما يكون الـ MSE بكامله ناشئًا عن الانحياز؟ هل تتفق نتائجك مع اتجاهات التغيّر الملحوظة في الجدول (١-٢) عندما تتغيّر قيمة B/σ من 0 إلى 0.6 ؟

(١ - ٨) عندما يكون ضروريًا مقارنة تقديرين يختلف فيهما التوزيعان التكراريان للخطأين $(\hat{\mu}-\mu)$. يمكن أحيانًا، وفي مسائل متخصصة، حساب الخسارة الناتجة عن خطأ $(\hat{\mu}-\mu)$ مهما كان حجمه. ويكون التقدير الأفضل هو التقدير الذي يؤدي إلى أقل قيمة لتوقع الخسارة، مع ثبات جميع العوامل الأخرى. بيّن أنّه إذا كانت الخسارة دالة تربيعية في الخطأ، أي من النوع $\lambda(\hat{\mu}-\mu)^2$ ، فينبغي اختيار التقدير الذي يحقق أصغر متوسط لمربع الخطأ.



المعاينة العشوائية البسيطة

(٢-١) المعاينة العشوائيّة البسيطة

المعاينة العشوائية البسيطة هي طريقة لاختيار n وحدة من بين N وحدة بحيث يكون لكل من العيّنات الـ ${}_NC_n$ الممكنة الفرصة نفسها في أن تكون هي العيّنة المسحوبة. وعمليًا تُسحب العيّنة العشوائية البسيطة وحدة فوحدة. ونُرقّم الوحدات في المجتمع من 1 إلى N . وعندئذ نسحب سلسلة من الأعداد العشوائية بين 1 و N ، إما بوساطة جدول للأعداد العشوائية أو بوساطة برنامج على الحاسب الآلي يُنتج مثل هذا الجدول. وعند كل سحب يجب أن تعطي الطريقة المستخدمة فرصة الاختيار نفسها لأي عدد من المجتمع لم يجرِ سحبه بعد. والوحدات التي تحمل هذه الأعداد الـ n تُشكّل العيّنة. ومن السهل التحقق من أن جميع العيّنات الـ ${}_NC_n$ المتميزة التي اختيرت بهذه الطريقة تتمتع بالفرصة نفسها. لنأخذ عيّنة متميزة واحدة، أي مجموعة من n وحدة محددة. فاحتمال اختيار واحدة من هذه الوحدات في السحب الأول هو n/N . وفي السحب الثاني نجد أن احتمال سحب إحدى الوحدات الـ n-1 الباقية هو $(n\text{-}1)/(N\text{-}1)$ ، وهلم جرًا. وبالتالي فإن احتمال اختيار تلك الوحدات المحددة الـ n خلال n سحبًا هو:

$$\frac{n}{N}\cdot\frac{(n-1)}{(N-1)}\cdot\frac{(n-2)}{(N-2)}\cdots\frac{1}{(N-n+1)}=\frac{n!(N-n)!}{(N)!}=\frac{1}{{}_NC_n} \tag{2.1}$$

وبما أننا نُخرج من المجتمع الرقم الذي يجري سحبه وذلك في كل عمليات السحب اللاحقة فتدعى هذه الطريقة أيضًا المعاينة بدون إعادة. والمعاينة العشوائية مع الإعادة في متناول اليد تمامًا: عند كل سحب يُعطى كل رقم من الأرقام الـ N في

المجتمع الفرصة نفسها في أن يكون هو الرقم المسحوب، وذلك بصرف النظر عن تكرار سحب أي رقم. والعلاقات الخاصة بالتباينات وتقدير تباينات التقديرات، التي نحسبها من العيّنة، غالبًا ما تكون في المعاينة مع الإعادة أبسط منها في المعاينة بدون إعادة. ولهذا السبب نستخدم أحيانًا المعاينة مع الإعادة في خطط المعاينة الأكثر تعقيدًا. وكما يبدو للوهلة الأولى فإن شمول العيّنة للوحدة نفسها مرتين أو أكثر أمر يفتقر إلى المنطق.

(٢-٢) اختيار عيّنة عشوائيّة بسيطة

جداول الأعداد العشوائية هي جداول من الأرقام ...,0,1,2 . وعند كل سحب، يكون لكل رقم من هذه الأرقام الفرصة نفسها في أن يكون الرقم المسحوب. ومن بين الجداول الأكثر انتشارًا نجد تلك التي نشرتها Rand Corporation (1955) - مليون رقم - وتلك التي نشرها Kendall & Smith (1938) -100,000 رقم. ويتوافر العديد من الجداول، كثير منها في الكتب المدرسية الإحصائية. ويعرض الجدول (٢-١) ألف رقم عشوائي للتوضيح، وهو من Snedecor & Cochran (1979) .

وعند استخدام هذه الجداول لاختيار عيّنة عشوائية بسيطة، فإن الخطوة الأولى هي ترقيم الوحدات في المجتمع من 1 إلى N . وإذا كان الرقم الأول من العدد N بين 5 و 9 ، تكون الطريقة التالية في الاختيار مناسبة. لنفرض $N=528$ ونريد $n=10$. نختار ثلاثة أعمدة من الجدول (٢-١)، ولتكن مثلاً الأعمدة من 25 إلى 27 . ولنستعرض أرقام الأعمدة الثلاثة من الأعلى إلى الأسفل ونختار الأعداد المتميزة العشرة الأولى بين 001 و 528 . فنجدها 36 ,509 ,364 ,417 ,348 ,127 ,149 ,186 ,290 ,162 . ومن أجل العددين الأخيرين قفزنا إلى الأعمدة 30 ,31 ,32 . وعند اختيار عيّنات مختلفة يستحسن تغيير النقطة التي نبدأ عندها في الجدول من عيّنة إلى أخرى.

ومساوىء هذه الطريقة هي أننا نستخدم الأعداد من ثلاثة أرقام 000 و 529 إلى 999 ، مع أن القفز عن أرقام لا يستهلك الكثير من الوقت. وعندما يكون الرقم الأول من العدد N أقل من 5 فقد يبقى البعض يفضل استخدام هذه الطريقة إذا كان n صغيرًا وكان جدول الأرقام العشوائية المتوافر كبيرًا.

جدول (٢-١) . ألف رقم عشوائي

	00–04	05–09	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49
00	54463	22662	65905	70639	79365	67382	29085	69831	47058	08186
01	15389	85205	18850	39226	42249	90669	96325	23248	60933	26927
02	85941	40756	82414	02015	13858	78030	16269	65978	01385	15345
03	61149	69440	11286	88218	58925	03638	52862	62733	33451	77455
04	05219	81619	10651	67079	92511	59888	84502	72095	83463	75577
05	41417	98326	87719	92294	46614	50948	64886	20002	97365	30976
06	28357	94070	20652	35774	16249	75019	21145	05217	47286	76305
07	17783	00015	10806	83091	91530	36466	39981	62481	49177	75779
08	40950	84820	29881	85966	62800	70326	84740	62660	77379	90279
09	82995	64157	66164	41180	10089	41757	78258	96488	88629	37231
10	96754	17676	55659	44105	47361	34833	86679	23930	53249	27083
11	34357	88040	53364	71726	45690	66334	60332	22554	90600	71113
12	06318	37403	49927	57715	50423	67372	63116	48888	21505	80182
13	62111	52820	07243	79931	89292	84767	85693	73947	22278	11551
14	47534	09243	67879	00544	23410	12740	02540	54440	32949	13491
15	98614	75993	84460	62846	59844	14922	48730	73443	48167	34770
16	24856	03648	44898	09351	98795	18644	39765	71058	90368	44104
17	96887	12479	80621	66223	86085	78285	02432	53342	42846	94771
18	90801	21472	42815	77408	37390	76766	52615	32141	30268	18106
19	55165	77312	83666	36028	28420	70219	81369	41943	47366	41067

ومن أجل $N=128$ ، مثلاً هناك طريقة ثانية سهلة التطبيق وفرص رفض الأرقام فيهـا أقل وهي كما يلي: في سلسلة الأعداد ذات الثلاثة أرقام، نطرح 200 من جميع الأعداد بين 201 و 400 ، ونطرح 400 من جميع الأعداد بين 401 و 600 ، ونطرح 600 من جميع الأعـداد بين 601 و 800 ، ونـطرح 800 من جميع الأعـداد بين 801 و 999 ، وبالطبع 000 من جميع الأعداد بين 000 و 200 . وكل باقٍ أكبر من 129 بالإضافة إلى الأعداد 000 ، 200 وهلم جرًّا ترفض. وباستخدام الأعمدة 05 إلى 07 في الجدول (٢-١) نجد 26 ,52 ,7 ,94 ,16 ,48 ,41 ,80 ,128 و 92 . ويتطلب السحب خمسة عشر عددًا ثلاثي الأرقـام في حالـة $n=10$. وفي هذه الـعـيّـنـة نجـد أن نسـبـة الـرفض $5/15=33\%$ قريبة من احتمال الرفض $72/200=36\%$ في هذه الطريقة. وعند استخدام هذه الـطريقـة في حالـة عدد N مثـل 384 ، نلاحظ أنـنـا نطرح 400 من عدد بين 401 و 800 ، إلا أنـنـا بصـورة آلية نرفض جميع الأعـداد التي هي أكـبر من

800 . وطـرح 800 من أعـداد بين 801 و 999 قد يعطي احتمال قبول أعلى لبواق بين 001 و 199 مما يعطيه لبواقٍ بين 200 و 384 .

وغالبًا ما نفضّل طرقًا أخرى على طريقة المعاينة العشوائية البسيطة وذلك على أساس من السهولة أو زيادة الإحكام . وأفضل ما تخدمه المعاينة العشوائية البسيطة هو أنها تُشكّل مدخلاً إلى نظرية المعاينة .

(٢-٣) تعاريـف ورمـوز

نتخـذ قرارنـا في مسـح عيّنة حول الخواص التي سنحاول، في كل وحدة من وحدات المعاينة التي تضمنتها العيّنة، قياسها ثم تسجيل نتيجة القياس . وسنشير إلى هذه الخواص كصفات مميزة، أو لمزيد من التبسيط، كمفردات .

ونرمز للقيم الموافقة لمفردة محددة وهي التي نحصل عليها من الوحدات الـ N التي تُشكّـل المجتمـع بالـرموز $y_1, y_2, \ldots, y_N$. ونرمز للقيم الموافقة بوحدات العيّنة بالرموز $y_1, y_2, \ldots, y_n$ ، أو إذا رغبنا في الإشارة إلى عنصر نموذجي من العيّنة فنكتب $y_i \; (i=1,2,\ldots,n)$

ولنلاحظ أن العيّنة سوف لا تتألف من الوحدات الـ n الأولى من المجتمع، باستثناء الحادثة التي يتفق أن نحصل عليها كنتيجة للسحب، وهيَ حادثة نادرة عادةً . وإذا تذكّرنا هذه الملاحظة باستمرار، فإن خبرتي تفيد بأنه سوف لا يكون هناك أي التباس .

وسنشـير إلى الصفـات المميزة لمجتمـع بأحرف كبيرة، وبأحرف صغيرة لتلك المتعلقة بالعينة . ولكل من المجاميع والمتوسطات لدينا التعاريف التالية :

	المجتمع	العيّنة
المجموع	$Y=\sum^{N} y_i = y_1+y_2+\cdots+y_N$	$\sum^{n} y_i = y_1+y_2+\cdots+y_n$
المتوسط	$\bar{Y}=\dfrac{y_1+y_2+\cdots+y_N}{N}=\dfrac{\sum^{N} y_i}{N}$	$\bar{y}=\dfrac{y_1+y_2+\cdots+y_n}{n}=\dfrac{\sum^{n} y_i}{n}$

ومع أننا نقوم بالمعاينة للعديد من الأهداف إلا أن الاهتمام يتركز في معظم الأحيان على أربع صفات مميزة للمجتمع.

١ - المتوسط = $\bar{Y}$ (مثلاً، معدّل عدد الأطفال في المدرسة الواحدة).
٢ - المجموع = Y (مثلاً، مجموع عدد فدادين القمح في المنطقة).
٣ - نسبة مجموعين أو متوسطين $R=Y/X=\bar{Y}/\bar{X}$
(مثلاً نسبة الممتلكات المنقولة إلى الممتلكات الكلية في مجموعة من الأسر).
٤ - نسبة الوحدات التي تقع ضمن صنف معين (مثلاً، نسبة الناس الذين يمتلكون أسنانًا صناعية).

ونناقش في هذا الفصل تقدير الكميات الثلاث الأولى. ويستخدم الرمز ^ للدلالة على تقدير حصلنا عليه من العيّنة لإحدى الصفات المميزة للمجتمع، وفي هذا الفصل سنأخذ بعين الاعتبار أبسط أنواع التقديرات.

	المقدِّر
$\bar{Y}$ متوسط المجتمع	متوسط العيّنة = $\hat{\bar{Y}} = \bar{y}$
Y المجموع الكلي للمجتمع	$\hat{Y} = N\bar{y} = N\sum^{n} y_i/n$
R النسبة في المجتمع	$\hat{R} = \bar{y}/\bar{x} = \sum^{n} y_i \Big/ \sum^{n} x_i$

وفي $\hat{Y}$ يدعى العامل N/n الذي نضرب به مجموع العيّنة أحيانًا، عامل التوسع، أو عامل النهوض، أو عامل التضخم. وعكسه n/N هو بالطبع نسبة حجم العيّنة إلى حجم المجتمع، ويدعى كسر المعاينة ونرمز له بالحرف f.

(٢-٤) خواص التقديرات

تعتمد دقة أي تقدير نقوم به من العيّنة على الطريقة التي نحسب فيها هذا التقدير من البيان الإحصائي للعيّنة، وعلى خطّة المعاينة. ونكتب أحيانًا للاختصار «دقة متوسط العيّنة» أو «دقة المعاينة العشوائية البسيطة». دون أن نذكر على وجه

التحديد العامل الأساسي الآخر. ونأمل أن يجري ذلك، فقط في أمثلة يتضح فيها من السياق ما هو العامل المحذوف. وعند دراسة أية علاقة نقدمها يجب أن يتأكد القارىء من أنه يعرف طريقة المعاينة وطريقة التقدير اللتين وضعت العلاقة من أجلهما.

وفي هذا الكتاب، سنقول إن طريقة المعاينة متّسقة إذا أصبح التقدير مساويًا تمامًا للقيمة المقدّرة من المجتمع وذلك عندما تصبح $N=n$ أي عندما تتألف العيّنة من المجتمع بكامله. ومن الواضح أنه في حالة المعاينة العشوائية البسيطة يكون $\bar{y}$ و $N\bar{y}$ تقديرين متّسقين لمتوسط المجتمع، ومجموع المجتمع على الترتيب. والاتّساق خاصة مرغوب فيها للتقديرات. وعلى الوجه الآخر فإن التقدير غير المتّسق، ليس عديم الفائدة بالضرورة، إذ يمكن أن يعطي دقة مُرضية في حالة كون n صغيرة بالمقارنة مع N. وعلى الأغلب فإن استخدامه مقصور على مثل هذه الحال. ويعطي Madow, Hurwitz & Hansen (1953) و Murthy (1976) تعريفًا بديلًا للاتّساق مماثلًا لذلك المعطى في الإحصاء التقليدي. فالتقدير يكون متسقًا إذا كان احتمال أن يتجاوز خطؤه أي قيمة معطاة ينتهي إلى الصفر عندما تصبح العيّنة كبيرة. والعبارة المضبوطة لهذا التعريف تحتاج إلى العناية والحذر في خطط المسوح الإحصائية المعقدة.

وكما رأينا فإن طريقة التقدير تكون غير منحازة إذا كان معدّل قيم التقدير محسوبًا فوق جميع العيّنات الممكنة ذات الحجم n، مساويًا تمامًا للقيمة الحقيقية الموافقة للمجتمع. ولكي تكون الطريقة غير منحازة فإن هذه النتيجة يجب أن تبقى صحيحة في أي مجتمع مقاديره، y_i منتهية، ولأي حجم عيّنة n، ولتقصّي ما إذا كان $\bar{y}$ غير منحاز في حالة المعاينة العشوائية البسيطة، نحسب قيمة $\bar{y}$ في جميع الـ ${}_NC_n$ من العيّنات الممكنة، ثم نحسب معدّل هذه القيم. ويدل الرمز E على عملية حساب المعدّل آخذين في الاعتبار جميع العيّنات الممكنة.

نظرية (٢-١)

متوسط العيّنة $\bar{y}$ هو تقدير غير منحاز لمتوسط المجتمع $\bar{Y}$.

برهان بالتعريف لدينا:

$$E\bar{y} = \frac{\sum \bar{y}}{{}_N C_n} = \frac{\sum (y_1 + y_2 + \cdots + y_n)}{n[N!/n!(N-n)!]} \tag{2.2}$$

حيث تمتد إشارة المجموع $\sum$ فوق مجمل الـ ${}_N C_n$ من العيّنات. ولحساب هذا المجموع نبحث عن عدد العيّنات التي تظهر فيها أي قيمة معيّنة y_i. وحيث إن هناك $(N-1)$ وحدة أخرى متاحة لاستكمال العيّنة وكذلك $(n-1)$ خانة أخرى من العيّنة ينبغي شغلها فإن عدد العيّنات التي تحوي y_i هو:

$$_{N-1}C_{n-1} = \frac{(N-1)!}{(n-1)!(N-n)!} \tag{2.3}$$

ومنه:

$$\sum (y_1 + y_2 + \cdots + y_n) = \frac{(N-1)!}{(n-1)!(N-n)!}(y_1 + y_2 + \cdots + y_N)$$

وبالاستناد إلى (2.2) فإن هذا يعطي:

$$E\bar{y} = \frac{(N-1)!}{(n-1)!(N-n)!} \frac{n!(N-n)!}{nN!}(y_1 + y_2 + \cdots + y_N)$$

$$= \frac{(y_1 + y_2 + \cdots + y_N)}{N} = \bar{Y} \tag{2.4}$$

نتيجة

$\hat{Y} = n\bar{y}$ هو تقدير غير منحاز لمجموع المجتمع Y .

وهناك برهان أقل تعقيدًا للنظرية (٢-١) نحصل عليه كما يلي: بما أن كل وحدة تظهر في عدد متساوٍ من العيّنات فمن الواضح أن:

(2.5) $E(y_1+y_2+...+y_n)$ يجب أن يكون مساويًا لجداء عدد بـ $(y_1+y_2+...+y_N)$

وعامل الضرب يجب أن يكون n/N باعتبار أن العبارة الأولى فيها n حدًا بينما تحوي العبارة الثانية N حدًا. ويقود هذا إلى النتيجة المطلوبة.

(٢-٥) تباينات التقديرات

نعرّف عادة تباين y_i في مجتمع منتهٍ بالعلاقة :

$$\sigma^2 = \frac{\sum_{1}^{N} (y_i - \bar{Y})^2}{N} \tag{2.6}$$

وسنصطلح على تقديم النتائج بدلالة عبارة مختلفة قليلاً يكون فيها المقام $(N-1)$ بدلاً من N . فنأخذ :

$$S^2 = \frac{\sum_{1}^{N} (y_i - \bar{Y})^2}{N-1} \tag{2.7}$$

ويستخدم هذا الشكل الاصطلاحي للتباين من قبل أولئك الذين يعالجون نظرية المعاينة مستخدمين تحليل التباين. وفائدته هي أن معظم النتائج تأخذ شكلاً أبسط بقليل. وتبقى النتائج نفسها عند استخدام أي من المصطلحين شريطة الثبات على استخدام الرمز نفسه.

وندرس الآن تباين $\bar{y}$. ونعني بهذا $E(\bar{y}-\bar{Y})^2$ محسوبًا من كل الـ ${}_NC_n$ من العيّنات الممكنة.

نظرية (٢-٢)

تباين المتوسط $\bar{y}$ لعيّنة عشوائية بسيطة هو:

$$V(\bar{y}) = E(\bar{y} - \bar{Y})^2 = \frac{S^2}{n}\frac{(N-n)}{N} = \frac{S^2}{n}(1-f) \tag{2.8}$$

حيث $f=n/N$ كسر المعاينة.

برهـان

$$n(\bar{y} - \bar{Y}) = (y_1 - \bar{Y}) + (y_2 - \bar{Y}) + \cdots + (y_n - \bar{Y}) \tag{2.9}$$

وبالاستناد إلى حجة التناظر ذاتها كما استخدمناها في العلاقة (2.5) نجد أن:

$$E[(y_1-\bar{Y})^2+\cdots+(y_n-\bar{Y})^2]=\frac{n}{N}[(y_1-\bar{Y})^2+\cdots+(y_N-\bar{Y})^2] \quad (2.10)$$

كما نجد أيضًا أن:

$$E[(y_1-\bar{Y})(y_2-\bar{Y})+(y_1-\bar{Y})(y_3-\bar{Y})+\cdots+(y_{n-1}-\bar{Y})(y_n-\bar{Y})]$$
$$=\frac{n(n-1)}{N(N-1)}[(y_1-\bar{Y})(y_2-\bar{Y})+(y_1-\bar{Y})(y_3-\bar{Y})$$
$$+\cdots+(y_{N-1}-\bar{Y})(y_N-\bar{Y})] \quad (2.11)$$

وفي المعادلة (2.11) تمتد مجاميع الجداءات فوق جميع الأزواج من الوحدات في العيّنة والمجتمع، على الترتيب. ويحوي المجموع على اليسار $n(n-1)/2$ من الحدود، وعلى اليمين $N(N-1)/2$ حدًا.

لنربع الآن (2.9) ولنحسب معدّلها فوق جميع العيّنات العشوائية البسيطة. وباستخدام (2.10) و (2.11) نحصل على:

$$n^2E(\bar{y}-\bar{Y})^2=\frac{n}{N}\left\{(y_1-\bar{Y})^2+\cdots+(y_N-\bar{Y})^2\right.$$
$$\left.+\frac{2(n-1)}{N-1}[(y_1-\bar{Y})(y_2-\bar{Y})+\cdots+(y_{N-1}-\bar{Y})(y_N-\bar{Y})]\right\}$$

وبإتمام المربعات بالنسبة لكل حد جدائي يمكن أن نكتب:

$$n^2E(\bar{y}-\bar{Y})^2=\frac{n}{N}\left\{\left(1-\frac{n-1}{N-1}\right)[(y_1-\bar{Y})^2+\cdots+(y_N-\bar{Y})^2]\right.$$
$$\left.+\frac{n-1}{N-1}[(y_1-\bar{Y})+\cdots+(y_N-\bar{Y})]^2\right\}$$

وينعدم الحد الثاني داخل القوس المربع باعتبار أن مجموع y_i يساوي $N\bar{Y}$. وبقسمة الطرفين على n^2 نجد:

$$V(\bar{y})=E(\bar{y}-\bar{Y})^2=\frac{N-n}{nN(N-1)}\sum_{i=1}^{N}(y_i-\bar{Y})^2=\frac{S^2}{n}\frac{(N-n)}{N}$$

وهو المطلوب.

نتيجة ١

الخطأ المعياري لـ $\bar{y}$ هو:

$$\sigma_{\bar{y}} = \frac{S}{\sqrt{n}}\sqrt{(N-n)/N} = \frac{S}{\sqrt{n}}\sqrt{1-f} \qquad (2.12)$$

نتيجة ٢

تباين $\hat{y}=N\bar{y}$، كتقدير لمجموع المجتمع Y هو:

$$V(\hat{Y}) = E(\hat{Y}-Y)^2 = \frac{N^2S^2}{n}\frac{(N-n)}{N} = \frac{N^2S^2}{n}(1-f) \qquad (2.13)$$

نتيجة ٣

الخطأ المعياري لـ $\hat{y}$ هو:

$$\sigma_{\hat{Y}} = \frac{NS}{\sqrt{n}}\sqrt{(N-n)/N} = \frac{NS}{\sqrt{n}}\sqrt{1-f} \qquad (2.14)$$

(٢-٦) التصحيح في حالة مجتمع منته

من المعروف جيدًا أن تباين متوسط عيّنة عشوائية حجمها n من مجتمع لانهائي هو σ^2/n . والتغيير الوحيد في هذه النتيجة عندما يكون المجتمع منتهيًا هو إدخال العامل الإضافي $(N-n)/N$. والعاملان $(N-n)/N$ في حالة التباين، و $\sqrt{(N-n)/N}$ في حالة الخطأ المعياري يُسمّيان عاملي التصحيح لمجتمع منته (ت م م). ويُكتبان بمقام $(N-1)$ بدلاً من N من قبل الكُتّاب الذين يقدمون النتائج بدلالة σ . وهذان العاملان قريبان من الواحد شريطة أن يبقى كسر المعاينة n/N منخفضًا، وهكذا فإنه لا يوجد لحجم المجتمع أي تأثير مباشر على الخطأ المعياري لمتوسط العيّنة. وعلى سبيل المثال، إذا كانت S نفسها في المجتمعين، فإن عيّنة حجمها 500 من مجتمع مؤلف من 200,000 تعطي تقديرًا لمتوسط المجتمع، دقته تقريبًا،

هي الدقة نفسها لتقدير عيّنة حجمها 500 من مجتمع من 10,000 . وغالبًا ما يجد الأشخاص غير الملمّين بالمعاينة صعوبة كبيرة في تصديق مثل هذه النتيجة، وفي الحقيقة تبدو هذه النتيجة مدهشة. وبالنسبة لهم يبدو واضحًا بالبداهة أنه إذا حصلنا على معلومات تتعلق بجزء صغير جدًا من المجتمع فقط، فإنه لا يمكن أن يكون متوسط العيّنة دقيقًا. ومن المفيد للقارىء أن يناقش لنفسه السبب الذي يجعل وجهة النظر هذه خاطئة.

وفي التطبيقات يمكن إهمال الـ(ت م م) حيثما لا يتجاوز كسر المعاينة 5 بالمائة، ولغايات عديدة، حتى إذا كان عاليًا حتى العشرة بالمائة. وتأثير تجاهل التصحيح هو المبالغة في تقدير الخطأ المعياري للتقدير $\bar{y}$.

ولا نحتاج إلى النظرية التالية، وهي تعميم للنظرية (٢-٢)، في المناقشات التي يحويها هذا الفصل، ولكننا نبرهنها هنا لاستخدامها فيما بعد.

نظرية (٢-٣)

إذا كان y_i ، x_i زوجًا من المتغيرات معرّفًا في كل وحدة في المجتمع، و $\bar{y}$ ، $\bar{x}$ هما المتوسطان الموافقان لعيّنة عشوائية بسيطة حجمها n ، فعندئذ يكون التغاير:

$$E(\bar{y}-\bar{Y})(\bar{x}-\bar{X})=\frac{N-n}{nN}\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})(x_i-\bar{X}) \qquad (2.15)$$

وتختزل هذه النظرية إلى تلك في (٢-٢) إذا بقيت قيم المتغيرين y_i ، x_i نفسها في كل وحدة.

برهان

لنطبق النظرية (٢-٢) على المتغير $u_i=y_i+x_i$. فمتوسط المجتمع لـ u_i هو $\bar{U}=\bar{y}+\bar{x}$. وتعطي النظرية (٢-٢):

$$E(\bar{u}-\bar{U})^2=\frac{N-n}{nN}\frac{1}{N-1}\sum_{i=1}^{N}(u_i-\bar{U})^2$$

أي أن

$$E[(\bar{y}-\bar{Y})+(\bar{x}-\bar{X})]^2=\frac{N-n}{nN}\frac{1}{N-1}\sum_{i=1}^{N}[(y_i-\bar{Y})+(x_i-\bar{X})]^2 \qquad (2.16)$$

بنشر المربّعين في كلي الطرفين، نجد بالاستناد إلى النظرية (٢-٢) أن:

$$E(\bar{y}-\bar{Y})^2=\frac{N-n}{nN}\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2$$

وهناك علاقة مماثلة لـ $E(\bar{x}-\bar{X})^2$. وهكذا تُحذف الحدود المتساوية في طرفي المعادلة (2.16) . والمطلوب في النظرية، أي المعادلة (2.15) ، ينتج من الحدود الجدائية في الطرفين.

(٢-٧) تقدير الخطأ المعياري من العيّنة

نستخدم عادةً، العلاقتين الموافقتين للخطأ المعياري لكل من تقدير متوسط مجتمع وتقدير مجموع مجتمع لغايات ثلاث هي: (1) مقارنة الدقة الناتجة عن المعاينة العشوائية البسيطة مع تلك الناتجة عن طرق أخرى في المعاينة، (2) تقدير حجم العيّنة الذي نحتاجه في مسح نقوم بتخطيطه و (3) تقدير الدقة الفعلية التي بلغناها في مسح تمّ إنجازه. والعلاقات تحوي تباين المجتمع S^2 ، وهو غير معروف في التطبيقات العملية، ولكن يمكن تقديره من البيانات الإحصائية في العيّنة. والنتيجة المناسبة معروضة في النظرية (٢-٤).

نظرية (٢-٤)

في حالة عيّنة عشوائية بسيطة يكون:

$$s^2=\frac{\sum_{1}^{n}(y_i-\bar{y})^2}{n-1}$$

تقديرًا غير منحاز لـ

$$S^2 = \frac{\sum_{1}^{N} (y_i - \bar{Y})^2}{N-1}$$

برهان

يمكن كتابة :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} [(y_i - \bar{Y}) - (\bar{y} - \bar{Y})]^2 \tag{2.17}$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^{n} (y_i - \bar{Y})^2 - n(\bar{y} - \bar{Y})^2 \right] \tag{2.18}$$

والآن لنأخذ المعدّل فوق جميع العيّنات العشوائية البسيطة التي حجمها n . وبالاستناد إلى حجة التناظر ذاتها المستخدمة في النظرية (٢-٢) نجد :

$$E\left[\sum_{i=1}^{n} (y_i - \bar{Y})^2\right] = \frac{n}{N} \sum_{i=1}^{N} (y_i - \bar{Y})^2 = \frac{n(N-1)}{N} S^2$$

وذلك من تعريف S^2 . وبالإضافة إلى ذلك، نجد من النظرية (٢-٢) :

$$E[n(\bar{y} - \bar{Y})^2] = \frac{N-n}{N} S^2$$

وبالتالي :

$$E(s^2) = \frac{S^2}{(n-1)N} [n(N-1) - (N-n)] = S^2 \tag{2.19}$$

نتيجة

التقديران غير المنحازين لتباينى $\bar{y}$ و $\hat{Y} = N\bar{y}$ هما :

$$v(\bar{y}) = s_{\bar{y}}^2 = \frac{s^2}{n}\left(\frac{N-n}{N}\right) = \frac{s^2}{n}(1-f) \tag{2.20}$$

$$v(\hat{Y}) = s_{\hat{Y}}^2 = \frac{N^2 s^2}{n}\left(\frac{N-n}{N}\right) = \frac{N^2 s^2}{n}(1-f) \tag{2.21}$$

ونأخذ في حالة الأخطاء المعيارية :

$$s_{\bar{y}} = \frac{s}{\sqrt{n}}\sqrt{1-f}, \qquad s_{\hat{Y}} = \frac{Ns}{\sqrt{n}}\sqrt{1-f} \tag{2.22}$$

وهذان التقديران منحازان بصورة طفيفة : وفي معظم التطبيقات العملية يكون الانحياز غير مهم .

وينبغي أن يلاحظ القارىء الرموز المستخدمة لتباينات التقديرات الفعلي منها والمقدّر وهكذا نكتب من أجل $\bar{y}$:

تباين فعلي	$V(\bar{y}) = \sigma_{\bar{y}}^2$
تباين مقدّر	$v(\bar{y}) = s_{\bar{y}}^2$

(٢ ـ ٨) حدود الثقة

نفرض عادة أن التقديرين $\bar{y}$ و $\hat{Y}$ يتوزعان طبيعيًا حول قيم المجتمع الموافقة . ونناقش في الفقرة (٢ـ١٥) أسباب وآفاق مثل هذا الفرض . وإذا كان هذا الفرض قائمًا فإن حدود الثقة الدنيا والعليا لمتوسط المجتمع ومجموعه تكون كما يلي :

المتوسط

$$\hat{\bar{Y}}_L = \bar{y} - \frac{ts}{\sqrt{n}}\sqrt{1-f}, \qquad \hat{\bar{Y}}_U = \bar{y} + \frac{ts}{\sqrt{n}}\sqrt{1-f} \tag{2.23}$$

المجموع :

$$\hat{Y}_L = N\bar{y} - \frac{tNs}{\sqrt{n}}\sqrt{1-f}, \qquad \hat{Y}_U = N\bar{y} + \frac{tNs}{\sqrt{n}}\sqrt{1-f} \tag{2.24}$$

حيث يرمز t لقيمة المتغير الطبيعي المعياري الموافقة لاحتمال الثقة المرغوب . والقيم الأكثر استخداما هي :

احتمال الثقة (%)	50	80	90	95	99
t	0.67	1.28	1.64	1.96	2.58

وإذا كان حجم العيّنة أقل من 50 فيمكن الحصول على قيم t الموافقة من جدول توزيع ستيودنت بـ (n−1) درجة من الحرية، وهي درجات الحرية الموافقة لتقدير

التبـاين s^2 أو يكـون استخدام التوزيع t مشروعًا تمامًا، فقط عندما تتبع الملاحظات y_i نفسها التوزيع الطبيعي، ويكون حجم المجتمع N لا نهائيًا. وليس للحيدان المعتدل عن التوزيع الطبيعي تأثير كبير في النتائج. ونحتاج إلى طرق خاصة في حالة عيّنات صغيرة وتوزيعات بعيدة عن التناظر (شديدة الالتواء).

مثـال

جُمعت تواقيع عريضة على 676 صفحة ورق. وكل صفحة تتسع لِـ 42 توقيعًا. ولكن عددًا أصغر من التواقيع جُمع على العديد من هذه الصفحات. وقد أحصينا عدد التواقيع في كل صفحة من صفحات عيّنة عشوائية حجمها $n=50$ (عيّنة مؤلفة من حوالي 7 بالمائة من المجتمع)، وكانت النتائج كما هو مبيّن في الجدول (٢-٢).

قدّر العدد الكلي للتواقيع في العريضة، وضَع 80% حدود ثقة لهذا التقدير.

وحدة المعاينة هي الصفحة، والملاحظات y_i هي أعداد التواقيع في كل صفحة. وبـما أن حوالي نصف عدد صفحـات العيّنـة يحوي العـدد الأعـظم من التـواقيع، أي 42، فقد قدمنا البيان الإحصائي على شكل جدول للتكرار. ونلاحظ أن التوزيع يبدو وكأنه في الأصل بعيد عن كونه طبيعيًا، فأكبر تكرار يقع من أجل القيمة الأكبر لِـ y_i. ومـع ذلـك فهناك أسباب للاعتقاد، استنادًا للخبرة العملية، بأن متوسطات العيّنات التي حجمها 50 تتوزع تقريبًا وفق التوزيع الطبيعي.

لدينا:

$$n = \sum f_i = 50, \qquad y = \sum f_i y_i = 1471, \qquad \sum f_i y_i^2 = 54{,}497$$

ومنه يكون تقدير مجموع عدد التواقيع:

$$\hat{Y} = N\bar{y} = \frac{(676)(1471)}{50} = 19{,}888$$

وبحساب تباين العيّنة s^2 نجد:

$$s^2 = \frac{1}{n-1}[\sum f_i (y_i - \bar{y})^2] = \frac{1}{n-1}\left[\sum f_i y_i^2 - \frac{(\sum f_i y_i)^2}{\sum f_i}\right]$$

$$= \frac{1}{49}\left[54{,}497 - \frac{(1471)^2}{50}\right] = 229.0$$

ومن المعادلة (2.24) نجد أن 80% حدود ثقة للمجموع هي :

$$19,888 \pm \frac{tNs}{\sqrt{n}}\sqrt{1-f} = 19,888 \pm \frac{(1.28)(676)(15.13)\sqrt{1-0.0740}}{\sqrt{50}}$$

وهذا يعطي 80% حدود ثقة 18107 و 21669 . وعند القيام بالعدّ الكامل تبيّن أن العدد الكلي هو في الحقيقة 21045 توقيعًا .

جدول (٢-٢) النتائج في عيّنة من 50 من صفحات العريضة: y_i عدد التواقيع ، f_i التكرار.

y_i	42	41	36	32	29	27	23	19	16	15
f_i	23	4	1	1	1	2	1	1	2	2
y_i	14	11	10	9	7	6	5	4	3	المجموع
f_i	1	1	1	1	1	3	2	1	1	50

(٢-٩) طريقة بديلة للبرهان

اقترح Cornfield (1944) طريقة لبرهان النتائج الرئيسة في المعاينة العشوائية البسيطة بدون إعادة، وهي طريقة تمكّننا من استخدام النتائج المتعارف عليها في نظرية المجتمعات اللانهائية. ليكن a_i متغيّرًا عشوائيًا يأخذ القيمة 1 إذا كانت الوحدة i ضمن العيّنة والقيمة 0 فيما عدا ذلك. فيمكن كتابة متوسط العيّنة $\bar{y}$ على الصيغة

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{N} a_i y_i \qquad (2.25)$$

حيث يمتد المجموع فوق جميع الوحدات الـ N في المجتمع. والمقادير a_i ، في هذه العبارة، هي متغيّرات عشوائية بينما الـ y_i هي مجموعة من الأعداد المثبتة.

ومن الواضح أن

$$\Pr(a_i = 1) = \frac{n}{N}, \qquad \Pr(a_i = 0) = 1 - \frac{n}{N}$$

وهكذا فإن a_i يتوزّع كمتغيّر ثنائي مع تكرار واحد، حيث $P = n/N$ ومنه

$$E(a_i) = P = \frac{n}{N}, \qquad V(a_i) = PQ = \frac{n}{N}\left(1 - \frac{n}{N}\right) \qquad (2.26)$$

ولإيجــاد $V(\bar{y})$ نحتــاج أيضًــا إلى تغــاير a_i و a_j . والجــداء a_i . a_j يســاوي 1 إذا كانت كل من الوحدة i والوحدة j ضمن العيّنة، ويساوي الصفر فيما عدا ذلك. ونـرى بسهـولـة أن احتمال أن تتضمن العيّنة وحدتين معينتين هو $n(n-1)/N(N-1)$.

وبالتالي

$$\mathrm{Cov}(a_i a_j) = E(a_i a_j) - E(a_i)E(a_j)$$

$$= \frac{n(n-1)}{N(N-1)} - \left(\frac{n}{N}\right)^2 = -\frac{n}{N(N-1)}\left(1-\frac{n}{N}\right) \tag{2.27}$$

وبتطبيق هذه الطريقة لإيجاد $V(\bar{y})$ نجد من (2.25)

$$V(\bar{y}) = \frac{1}{n^2}\left[\sum_{i=1}^{N} y_i^2 V(a_i) + 2\sum_{i<j}^{N} y_i y_j \,\mathrm{Cov}\,(a_i a_j)\right] \tag{2.28}$$

$$= \frac{1-f}{nN}\left(\sum y_i^2 - \frac{2}{N-1}\sum y_i y_j\right) \tag{2.29}$$

وذلك باستخدام (2.26) و (2.27) . وبالإتمام إلى مربع كامل نجد أن

$$V(\bar{y}) = \frac{1-f}{nN}\left(\frac{N}{N-1}\sum y_i^2 - \frac{1}{N-1}Y^2\right) \tag{2.30}$$

$$= \frac{1-f}{n(N-1)}\sum (y_i - \bar{Y})^2 = \frac{(1-f)S^2}{n} \tag{2.31}$$

وتقدم هذه الطريقة براهين سهلة للنظريتين (2.3) و (2.4) . ويمكن استخدامها لإيجاد عزوم أعــلى لتــوزيع $\bar{y}$ ، وهنــاك طريقــة أقـوى بالنسبـة لهذا الغـرض، توصل إليها Tukey (1950) مع تطوير إضافي قام به Wishart (1952) .

(٢-١٠) المعاينة العشوائية مع الإعادة

وينطبق أسلوب مشابه عندما تكون المعاينة مع الإعادة. وفي هذه الحالة يمكن أن تظهـر الـوحـدة i في العيّنـة $0,1,2,\ldots,n$ مرة. ليكن t_i عدد المرات الذي تظهر فيه الوحدة i في العيّنة. فعندئذ:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{N} t_i y_i \tag{2.32}$$

وبما أنّ احتمال سحب الوحدة i هو $1/N$ وذلك عند كل سحب فالمتغيّر t_i يتوزع وفق التوزيع الثنائي بعدد من التكرارات يساوي n واحتمال نجاح $p=1/N$. وبالتالي

$$E(t_i)=\frac{n}{N}, \qquad V(t_i)=n\left(\frac{1}{N}\right)\left(1-\frac{1}{N}\right) \tag{2.33}$$

وبصورة مشتركة تتبع المتغيّرات t_i التوزيع متعدد الحدود. وفي هذا التوزيع :

$$\text{Cov}\,(t_i t_j)=-\frac{n}{N^2} \tag{2.34}$$

وباستخدام (2.32)، (2.33) ، نجد في حالة معاينة مع الإعادة :

$$V(\bar{y})=\frac{1}{n^2}\left[\sum_{i=1}^{N} y_i^2\frac{n(N-1)}{N^2}-2\sum_{i<j}^{N} y_i y_j\frac{n}{N^2}\right] \tag{2.35}$$

$$=\frac{1}{nN}\sum_{i=1}^{N}(y_i-\bar{Y})^2=\frac{\sigma^2}{n}=\frac{N-1}{N}\frac{S^2}{n} \tag{2.36}$$

ومنه يكون $V(\bar{y})$ في المعاينة بدون إعادة $(N-n)/(N-1)$ مرة فقط من قيمته عند المعاينة مع الإعادة. وإذا استخدمنا بدلاً من $\bar{y}$ المتوسط $\bar{y}_d$ وهو متوسط الوحدات المتميّزة أو المختلفة في العينة كتقدير عند المعاينة مع الإعادة. فقد بيّن Murthy (1967) أن الحد الرئيسي في التباين المتوسط لـ $\bar{y}_d$ هو $(1-\frac{f}{2})S^2/n$ مقتفيًا أثر بحوث قام بها Khamis, DesRaj (1958) و Basu (1958) . وفي بعض التطبيقات نجد أن تكلفة قياس الوحدات المتميّزة في العيّنة هي المهيمنة بحيث تصبح تكلفة العيّنة متناسقة مع عدد الوحدات المتميزة. وفي هذه الحالة يبيّن Seth (1964) و J.N.K. Rao (1964) أنه من أجل معدّل تكلفة معطى يكون $V(\bar{y})$ في المعاينة مع الإعادة. كما برهنا النتيجة الأعم وهي أنــه إذا كان $\bar{y}_{d'}=f(\nu)\bar{y}_d/Ef(\nu)$ ، حيث ν هو عدد الوحدات المتميزة في العيّنة و $f(\nu)$ دالّة في ν ، فعندئذ يكون $V(\bar{y})<V(\bar{y}_{d'})$ (شريطة أن يكون $S^2<N_{\bar{y}}$. وهذا الشرط محقق تقريبًا في جميع المجتمعات التي نواجهها في المسوح الإحصائية.

(٢-١١) تقدير نسبة

كثيرًا ما تكون الكمية التي نريد تقديرها من عيّنة عشوائية بسيطة هي نسبة متغيّرين يتغير كلاهما من وحدة إلى أخرى. ففي المسوح الإحصائية المنزلية نذكر كأمثلة متوسط عدد البذّات لكل ذكر بالغ، متوسط مصاريف أدوات التجميل لكل أنثى بالغة، ومتوسط عدد الساعات الأسبوعية المنقضية في مشاهدة التلفاز لكل طفل بين العاشرة والخامسة عشرة من العمر. ولكي نقدّر المفردة الأولى من هذه المفردات، يمكن أن نسجّل من أجل المنزل i $(i=1,2,\ldots,n)$ عدد البالغين الذكور x_i الذين يعيشون فيه والعدد الكلي من البذّات التي يملكونها y_i. ويكون متوسط المجتمع الذي نريد تقديره هو

$$R=\frac{\text{العدد الكلي من البذّات}}{\text{العدد الكلي من الذكور البالغين}}=\frac{\sum_1^N y_i}{\sum_1^N x_i} \tag{2.37}$$

وتقدير العيّنة الموافق هو

$$\hat{R}=\frac{\sum_1^n y_i}{\sum_1^n x_i}=\frac{\bar{y}}{\bar{x}} \tag{2.38}$$

وكثيرًا ما تقع أمثلة من هذا النوع عندما تتألف وحدة المعاينة (المنزل) من مجموعة أو جملة من العناصر (الذكور البالغون)، وينصبّ اهتمامنا في المجتمع على المتوسط لكل عنصر. وتظهر النسب أيضًا في العديد من التطبيقات الأخرى، فمثلًا، نسبة قروض البناء إلى مجموع القروض في مصرف، أو نسبة فدادين القمح إلى العدد الكلي من الفدادين في مزرعة.

والتوزيع الاحتمالي لـ $\hat{R}$ هو أكثر تعقيدًا من توزيع $\bar{y}$ بسبب تغير كل من البسط $\bar{y}$ والمقام $\bar{x}$ من عيّنة إلى عيّنة. وفي العيّنات الصغيرة يكون توزيع $\hat{R}$ غير متناظر وتشكل $\hat{R}$ عادةً تقديرًا منحازًا بصورة طفيفة لـ R . وفي العيّنات الكبيرة ينزع

توزيع $\hat{R}$ إلى أن يكون طبيعيًا ويصبح الانحياز مهملاً. ونستخدم النتيجة التقريبية التالية في معظم التطبيقات: ندرس توزيع $\hat{R}$ بتفصيل أكبر في الفصل السادس.

نظريـة (٢ـ٥)

إذا قيس y ، x في كل وحدة من عيّنة عشوائية بسيطة حجمها n ، نفترضه كبيراً فيكـون كل من متـوسط مربعـات الخطأ MSE الموافق لِـ $\hat{R}$ وتباين $\hat{R}$ معطى تقريبًا بالعلاقة:

$$\text{MSE}(\hat{R}) \doteq V(\hat{R}) \doteq \frac{1-f}{n\bar{X}^2} \frac{\sum_{i=1}^{N} (y_i - Rx_i)^2}{N-1} \tag{2.39}$$

حيث $R=\bar{Y}/\bar{X}$ هي نسبة متوسطي المجتمع و $f=n/N$.

برهان

$$\hat{R} - R = \frac{\bar{y}}{\bar{x}} - R = \frac{\bar{y} - R\bar{x}}{\bar{x}} \tag{2.40}$$

وإذا كان n كبيراً فينبغي ألا يختلف $\bar{x}$ كثيراً عن $\bar{X}$. ولتجنّب الغوص في توزيع نسبة متغـيّرين عشـوائيين $\bar{y}-R\bar{x}$ و $\bar{x}$ ، نضـع، كتقـريب، $\bar{X}$ بدلاً من $\bar{x}$ في المقـام في (2.40) وهذا يعطي:

$$\hat{R} - R \doteq \frac{\bar{y} - R\bar{x}}{\bar{X}} \tag{2.41}$$

والآن لنأخذ المتوسط فوق جميع العيّنات العشوائية ذات الحجم n .

$$E(\hat{R} - R) \doteq \frac{E(\bar{y} - R\bar{x})}{\bar{X}} = \frac{\bar{Y} - R\bar{X}}{\bar{X}} = 0 \tag{2.42}$$

باعتبار أن $R=\bar{Y}/\bar{X}$. ويبيّن هذا أن $\hat{R}$ تقدير غير منحاز لِـ R وذلك في حدود مرتبة التقريب المستخدم هنا.

ومن (2.41) نحصل أيضًا على النتيجة التالية:

$$\text{MSE}(\hat{R}) = E(\hat{R} - R)^2 \doteq \frac{1}{\bar{X}^2} E(\bar{y} - R\bar{x})^2 \qquad (2.43)$$

والمقدار $\bar{y} - R\bar{x}$ هو متوسط العيّنة للمتغير $d_i = y_i - Rx_i$ الذي متوسط مجتمعه هو $\bar{D} = \bar{y} - R\bar{\bar{x}} = 0$. وبـالتـالي يمكن إيجاد $V(\hat{R})$ بتطبيق النظرية (٢-٢) المتعلقة بتباين متوسط عيّنة عشوائية بسيطة، على المتغير d_i ثم التقسيم على $\bar{X}^2$. وهذا يعطي

$$V(\hat{R}) \doteq \frac{1}{\bar{X}^2} E(\bar{y} - R\bar{x})^2 = \frac{1}{\bar{X}^2} \frac{S_d^2}{n}(1-f) \qquad (2.44)$$

$$= \frac{1-f}{n\bar{X}^2} \frac{\sum_{i=1}^{N} (d_i - \bar{D})^2}{(N-1)} = \frac{1-f}{n\bar{X}^2} \frac{\sum_{i=1}^{N} (y_i - Rx_i)^2}{N-1} \qquad (2.45)$$

وهو المطلوب.

والطريقة التي بُرهنت فيها النظرية (٢-٥) تستحق التأمل. وقد بُرهن على أن العلاقة في النظرية (٢-٢) الخاصة بتباين متوسط عينة $\bar{y}$ تعطي العلاقة الخاصة بالتباين التقريبي للنسبة $\bar{y}/\bar{x}$ ، إذا أخذنا المتغيّر $(y_i - Rx_i)/\bar{X}$ بدلاً من المتغير y_i وتصحّ النتيجة نفسها، أو امتدادها الطبيعي، أيضًا في حالات معاينة أكثر تعقيدًا. وسنستخدم ذلك بصورة متكررة فيما بعد عبر هذا الكتاب.

كتقدير عيّنة لِـ

$$\frac{\sum_{i=1}^{N} (y_i - Rx_i)^2}{N-1}$$

من المعتاد أخذ:

$$\frac{\sum_{i=1}^{n} (y_i - \hat{R}x_i)^2}{n-1}$$

ويمكن البرهان على أن لهذا التقدير انحيازًا من مرتبة $1/n$.

وعند تقدير الخطأ المعياري لِـ $\hat{R}$ ، يُعطي هذا

$$s(\hat{R}) = \frac{\sqrt{1-f}}{\sqrt{n}\bar{X}} \sqrt{\frac{\sum (y_i - \hat{R}x_i)^2}{n-1}} \tag{2.46}$$

وإذا كان $\bar{X}$ غير معروف، نعوّض تقدير العيّنة $\bar{x}$ بدلاً عنه في المقام.
وإحدى الطرق لحساب $s(\hat{R})$ هو التعبير عنه على الشكل

$$s(\hat{R}) = \frac{\sqrt{1-f}}{\sqrt{n}\bar{X}} \sqrt{\frac{\sum y_i^2 - 2\hat{R} \sum y_i x_i + \hat{R}^2 \sum x_i^2}{n-1}} \tag{2.47}$$

مثال

يبيّن الجدول (٢-٣)، عدد الأشخاص (x_1) ، والدخل الأسبوعي للأسرة (x_2) ، والمصروف الأسبوعي على الطعام (y) ، في عيّنة عشوائية بسيطة من 33 أسرة من ذوي الدخل المحدود. وبما أن العيّنة صغيرة، فيهدف البيان فقط إلى توضيح الحسابات.

قدّر من العيّنة (ا) متوسط المصروف الأسبوعي على الطعام للأسرة الواحدة، (ب) متوسط مصروف الطعام الأسبوعي للشخص الواحد، و(ج) النسبة المئوية من الدخل المصروفة على الطعام. احسب الأخطاء المعيارية لهذه التقديرات.

مصروفات الطعام الأسبوعية للأسرة الواحدة: وهذا هو المتوسط العادي للعيّنة

$$\bar{y} = \frac{907.2}{33} = \$27.49$$

ومن النظرية (٢-٢) [متجاهلين الـ (ت م م)]، نجد أن خطأه المعياري

$$s_{\bar{y}} = \frac{1}{\sqrt{n}} \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = \frac{1}{\sqrt{n(n-1)}} \sqrt{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}$$

$$= \frac{1}{\sqrt{(33)(32)}} \sqrt{28{,}224 - (907.2)^2/33} = \$1.76$$

[مجموع المربعات غير المصحح 28224 معطى في أسفل الجدول (٢-٣)]

مصروفات الطعام الأسبوعية للشخص الواحد : بما أن حجم الأسرة متغير فالتقدير هو نسبة متغيرين

للشخص الواحد $\hat{R}_1 = \frac{\sum y}{\sum x_1} = \frac{907.2}{123} = \$7.38$

ومجاميع المربعات والجداءات التي نحتاجها لحساب $S(\hat{R})$ ، والمذكور في (2.47) نجدها تحت الجدول (٢-٣) . وبالإضافة إلى ذلك نحتاج إلى :

$$2\hat{R}_1 = 14.7512, \qquad \hat{R}_1{}^2 = 54.3996, \qquad \bar{x}_1 = 3.7273$$

وقد حملنا أرقاماً عشرية إضافية في $\hat{R}_1$ ، $2\hat{R}_1$ ، $\hat{R}_1^2$ للمحافظة على الدقة المطلوبة .

وهكذا نجد من (2.47)

$$s(\hat{R}_1) = \frac{1}{\sqrt{33}(3.7273)}\sqrt{\frac{(28,224) - (14.7512)(3595.5) + (54.3996)(533)}{32}}$$
$$= \$0.534$$

نسبة الدخل المصروفة على الطعام : من جديد نجد هنا نسبة متغيّرين

$$\hat{R}_2 = 100\frac{\sum y}{\sum x_2} = \frac{(100)(907.2)}{2394} = 37.9\%$$

ومن (2.47) يمكن للقارىء أن يتحقق من أن الخطأ المعياري هو 2.38%

(٢-١٢) تقديرات المتوسطات فوق مجتمعات جزئيّة

في العديد من المسوح الإحصائية نقوم بالتقديرات لكل صف من عدد من الصفوف التي نقسّم إليها المجتمع . وفي مسح منزلي قد نرغب بتقديرات تتعلق بأسر لديها ...,0,1,2 من الأطفال، أو تتعلق بالمالكين والمستأجرين، أو بأسر في مجموعات مهنية مختلفة . وقد اصطلحت اللجنة الفرعية للمعاينة التابعة للأمم المتحدة (1950) على تسمية مثل هذه المجتمعات

جدول (٢-٣) الحجم، الدخل الأسبوعي، وتكلفة الطعام لـ 33 أسرة

رقم الأسرة	الحجم x_1	الدخل x_2	كلفة الطعام y
1	2	62	14.3
2	3	62	20.8
3	3	87	22.7
4	5	65	30.5
5	4	58	41.2
6	7	92	28.2
7	2	88	24.2
8	4	79	30.0
9	2	83	24.2
10	5	62	44.4
11	3	63	13.4
12	6	62	19.8
13	4	60	29.4
14	4	75	27.1
15	2	90	22.2
16	5	75	37.7
17	3	69	22.6
18	4	83	36.0
19	2	85	20.6
20	4	73	27.7
21	2	66	25.9
22	5	58	23.3
23	3	77	39.8
24	4	69	16.8
25	7	65	37.8
26	3	77	34.8
27	3	69	28.7
28	6	95	63.0
29	2	77	19.5
30	2	69	21.6
31	6	69	18.2
32	4	67	20.1
33	2	63	20.7
Total	123	2394	907.2

$\sum x_1^2 = 533,\quad \sum x_2^2 = 177{,}254,\quad \sum y^2 = 28{,}224$

$\sum x_1 y = 3595.5,\quad \sum x_2 y = 66{,}678$

الجزئية بميادين *الدراسة*. وفي أبسط الحالات تقع كل وحدة في المجتمع في أحد الميادين. لنفرض أن الميدان j يتضمن N_j من الوحدات، وأن n_j هو عدد الوحدات في عيّنة عشوائية بسيطة حجمها n اتفق أن وقعت جميعها في هذا الميدان. وإذا كانت y_{jk} $(k=1,2,\ldots,n)$ هي القياسات في هذه الوحدات، فنقدّر متوسط المجتمع $\bar{y}_j$ في الميدان j بـ

$$\bar{y}_j = \sum_{k=1}^{n_j} \frac{y_{jk}}{n_j} \tag{2.48}$$

وللوهلة الأولى تبدو $\bar{y}_j$ وكأنها تقدير لنسبة كما في الفقرة (٢-١١) فمع أن n مثبّت سيتغير n_j من عيّنة حجمها n إلى عيّنة أخرى. ويمكن تجنّب تعقيدات تقدير نسبة بأن نأخذ في الاعتبار توزيع $\bar{y}_j$ فوق عيّنات نثبّت فيها كلاً من n و n_j. ونفترض أن $n_j>0$.

وفي جملة العيّنات التي حددنا فيها n و n_j نجد أن احتمال سحب أي مجموعة محددة من n_j وحدة من الوحدات الـ N_j الموجودة في الميدان j هو

$$\frac{{}_{N-N_j}C_{n-n_j}}{{}_{N-N_j}C_{n-n_j} \cdot {}_{N_j}C_{n_j}} = \frac{1}{{}_{N_j}C_{n_j}}$$

وبما أن كل مجموعة محددة من n_j وحدة من وحدات الميدان يمكن أن تظهر مع جميع اختيارات $(n-n_j)$ وحدة من الوحدات الـ $(N-n_j)$ التي لا تقع في الميدان j ، فالبسط أعلاه هو عدد العيّنات التي تتضمن مجموعة محدّدة من n_j وحدة، والمقام هو العدد الكلي للعيّنات. ونستنتج أنه يمكن تطبيق النظريات (٢-١)، (٢-٢) و(٢-٤) على y_{jk} إذا وضعنا n_j بدلاً من n و N_j بدلاً من N .

(2.49) ومن النظرية (٢-١): $\bar{y}_j$ تقدير غير منحاز لـ $\bar{Y}_j$

(2.50) من النظرية (٢-٢): الخطأ المعياري لـ $\bar{y}_j$ هو $\frac{S_j}{\sqrt{n_j}}\sqrt{1-(n_j/N_j)}$

حيث

$$S_j^2 = \sum_{k=1}^{N_j} \frac{(y_{jk} - \bar{Y}_j)^2}{N_j - 1} \tag{2.51}$$

ومن النظرية (٢-٤): تقدير الخطأ المعياري لـ $\bar{y}_j$ هو

$$\frac{s_j}{\sqrt{n_j}}\sqrt{1-(n_j/N_j)} \tag{2.52}$$

حيث

$$s_j^2 = \sum_{k=1}^{n_j} \frac{(y_{jk} - \bar{y}_j)^2}{n_j - 1} \tag{2.53}$$

وإذا كانت قيمة N_j غير معروفة فيمكن استخدام المقدار n/N بدلاً من n_j/N_j عند حساب الـ ت م م. (في معاينة عشوائية بسيطة n_j/N_j هو تقدير غير منحاز لـ (n/N) .

(٢-١٣) تقديرات المجاميع في المجتمعات الجزئية

في قائمة الحسابات المستحقة في شركة. والتي دُفع بعضها ولم يُدفع البعض الآخر، قد نرغب في أخذ عيّنة لتقدير المبلغ الكلي للفواتير غير المدفوعة. وإذا كان N_j (عدد الفواتير غير المدفوعة في المجتمع) معروفًا، فلا توجد مشكلة. إذ أن تقدير العيّنة هو $N_j \bar{y}_j$ وخطؤه المعياري الشرطي هو N_j مضروبًا بالعبارة (2.50).

وبصورة بديلة، إذا كان المبلغ الإجمالي المستحق في القائمة معروفًا، فيمكن استخدام تقدير نسبة. وتعطي العيّنة تقديرًا لنسبة المبلغ الإجمالي للفواتير غير المدفوعة إلى قيمة جميع الفواتير. ويُضرب هذا بالمبلغ الإجمالي المستحق في القائمة وقد افترضنا أنه معروف. وإذا لم يكن لا N_j ولا الاستحقاقات الإجمالية معروفًا فلا يمكن القيام بهذه التقديرات. وبدلًا من ذلك، نضرب مجموع المقادير y في العيّنة مأخوذة من الوحدات الواقعة في الميدان j بعامل النهوض N/n . وهذا يعطي التقدير

$$\hat{Y}_j = \frac{N}{n} \sum_{k=1}^{n_j} y_{jk} \tag{2.54}$$

سنبيّن أن $\hat{Y}_j$ غير منحاز ونحصل على خطئه المعياري فوق عيّنات مكررة حجمها n . ولا تساعدنا في هذه المسألة حيلة الاحتفاظ بـ n_j مثبة شأنها شأن n . وعند تقديم البرهان نعود إلى الرموز الأصلية، حيث y_i القياس الذي نحصل عليه من الوحدة i في المجتمع ونعرّف لكل وحدة في المجتمع متغيّرًا جديدًا y'_i ، حيث

$$y_i' = \begin{cases} y_i & \text{إذا كانت الوحدة في الميدان } j \\ 0 & \text{فيما عدا ذلك} \end{cases}$$

ومجموع القياسات y'_i في المجتمع هو

$$\sum_{i=1}^{N} y_i' = \sum_{j\text{th dom}} y_i = Y_j \tag{2.55}$$

وفي عيّنـة عشـوائية بسيطة حجمهـا n ، يكـون $y_i'=y_i$ لكـل من الـوحـدات الـ n_i التي تقـع في الميدان j و $y_i'=0$ لكـل من الـوحدات الـ $n-n_j$ الباقية . وإذا كان $\bar{y}'$ متوسط العيّنة العادي للمقادير y_i' ، فعندئذ

$$N\bar{y}'=\frac{N}{n}\sum_{i=1}^{n}y_i'=\frac{N}{n}\sum_{k=1}^{n_j}y_{jk}=\hat{Y}_j \tag{2.56}$$

وتبـيـّن هذه النتيجـة أن التقـدير $\hat{Y}_j$ كما عرفنـاه في المعـادلـة (2.54) يسـاوي N مرة متوسط العيّنة للمقادير y_i' .

وفي عيّنـات متكـررة حجمهـا n يمكننـا بوضـوح تطبيق النـظريات (2.1) ، (2.2) ، و (2.4) على المتغيّرات y_i' . وهذه النظريات تبيّن أن $\hat{Y}_j$ هو تقدير غير منحاز لِـ Y_j بخطأ معياري

$$\sigma(\hat{Y}_j)=\frac{NS'}{\sqrt{n}}\sqrt{1-(n/N)} \tag{2.57}$$

حيث S' هو بالنسبـة للمقـادير y_i' الانحـراف المعياري للمجتمـع . ولكي نحسب S' ، نعتـبر المجتمع وكأنه مؤلف من الـ N_j من المقادير y_j الموجودة في الميدان j ومن $N-N_j$ من القيم المساوية للصفر . وهكذا يكون

$$S'^2=\frac{1}{N-1}\left(\sum_{j\text{th dom}}y_i^2-\frac{Y_j^2}{N}\right) \tag{2.58}$$

ومن النظرية (٢ ـ ٤) نجد أن تقدير عيّنة للخطأ المعياري لِـ $\hat{Y}_j$ هو

$$s(\hat{Y}_j)=\frac{Ns'}{\sqrt{n}}\sqrt{1-(n/N)} \tag{2.59}$$

وعنـد حساب s' ، تأخذ كل وحدة خارج الميدان j القيمة صفر . ويبدو لدى بعض الطلاب شعور نفسي مضاد إزاء القيام بذلك ، إلا أن الطريقة سليمة . وطرق هذه الفقرة والفقرة السابقة تنطبق أيضًا على مسوح يتضمن الإطار المستخدم فيها وحدات لا تنتمي إلى المجتمع كما جرى تعريفه ونوضح هذا التطبيق بمثال .

مثال

من قائمة من 2422 من المصاريف المنزلية الثانوية سحبنا عيّنة عشوائية بسيطة من 180 مفردة لتقدير المصروف الكلي للمنزل. ولم تعتبر أنواع معيّنة من المصاريف مناسبة (مصاريف تتعلق بالثياب وصيانة السيارة). ومن بين المفردات الـ 180 في العيّنة كان 152 منها مناسبًا. وكان المجموع ومجموع المربعات غير المصححة لتكاليف المواد المناسبة (بالدولار) كما يلي

$$\sum y_i' = 343.5, \qquad \sum y_i'^2 = 1491.38$$

والمطلوب تقدير التكلفة الإجمالية لمصاريف المنزل وإعطاء الخطأ المعياري للتقدير.

$$\hat{Y}_i = \frac{N}{n}\sum_{i=1}^{n} y_i' = \frac{(2422)(343.5)}{180} = \$4622$$

ومن (2.59) ،

$$s(\hat{Y}_i) = \frac{Ns'}{\sqrt{n}}\sqrt{1-(n/N)}$$

وعند حساب s' نعتبر عيّنتنا التي تحتوي 180 مفردة وكأنها تتضمن 28 صفرًا. وبالتالي

$$s'^2 = \frac{1}{(179)}\left[\sum y_i'^2 - \frac{(\sum y_i')^2}{180}\right]$$
$$= \frac{1}{(179)}\left[1491.38 - \frac{(343.5)^2}{180}\right] = 4.670$$

وأخيرًا

$$s_{\hat{Y}_i} = (2422)\sqrt{\frac{4.670}{180}\left(1-\frac{180}{2422}\right)} = \$375$$

والتقدير ليس دقيقًا، فمعامل اختلافه يساوي 375/4622 أي حوالي 8% .

وفي هذا المثال استُثنيت مصاريف صيانة السيارة والملابس كمصاريف غير مناسبة، وبالتالي اعتُبرت أرقامها في العيّنة أصفارًا. وفي بعض التطبيقات يكون من المعروف سلفًا أن بعض الوحدات في المجتمع لا تُسهم بشيء في المجموع المراد تقديره.

وعلى سبيل المثال، مخازن لا تبيع الحقائب في مسح إحصائي للمخازن يهدف إلى تقدير المبيعات الإجمالية للحقائب؛ وبعض وحدات المعاينة، على شكل مساحة، في دراسات تتعلق بالمزارع قد لا تحوي مزارع. ومن الممكن أحيانًا، وبمزيد من الجهود، أن نتعرّف على الـوحـدات التي لا تسهم بشيء ونعـدُّهـا، وهكـذا يكون $(N-N_j)$ في رموزنا، وبالتالي N_j معروفًا. وكنتيجة فإن مقدار انخفاض $V(\hat{Y}_j)$، عندما يكون N_j معروفًا، أمر يستحق الدراسة. وإذا كان N_j غير معروف فإن (2.57) تعطي:

$$V(\hat{Y}_j) = \frac{N^2 S'^2}{n}\left(1 - \frac{n}{N}\right)$$

وإذا كان $\bar{Y}_j$ و S_j المتوسط والانحراف المعياري في الميدان الذي نهتم به (أي بين الوحدات التي لا تساوي صفرًا) فيمكن للقارىء التحقق من أن

$$(N-1)S'^2 = (N_j - 1)S_j^2 + N_j \bar{Y}_j^2\left(1 - \frac{N_j}{N}\right) \tag{2.60}$$

وبما أن الحدود التي تتضمن $1/N_j$ و $1/N$ هي على الدوام مهملة تقريبًا فلدينا

$$S'^2 \doteq P_j S_j^2 + P_j Q_j \bar{Y}_j^2 \tag{2.61}$$

حيث $P_j = N_j/N$ و $Q_j = 1 - P_j$. وهذا يعطي

$$V(\hat{Y}_j) \doteq \frac{N^2}{n}(P_j S_j^2 + P_j Q_j \bar{Y}_j^2)\left(1 - \frac{n}{N}\right) \tag{2.62}$$

وإذا تعـرّفنا على الوحدات التي لا تساوي صفرًا، نسحب منها عيّنة حجمها n_j. ويكون تقدير مجموع الميدان $N_j \bar{y}_j$ بتباين

$$V(N_j \bar{y}_j) = \frac{N_j^2}{n_j} S_j^2\left(1 - \frac{n_j}{N_j}\right) = \frac{N^2}{n_j} P_j^2 S_j^2\left(1 - \frac{n_j}{N_j}\right) \tag{2.63}$$

والتباينات القابلة للمقارنة هي (2.62) و (2.63) وفي (2.62) العدد المتوسط للوحدات غير المساوية للصفر في عيّنة حجمها n هو nP_j. وإذا أخذنا $n_j = nP_j$ في (2.63)، بحيث إن

عدد الوحدات غير المساوية للصفر التي سنقيسها في كلتا الطريقتين هو نفسه تقريبًا، فتصبح (2.63) على الشكل

$$V(N_j\bar{y}_j) = \frac{N^2}{n} P_j S_j^2 \left(1 - \frac{n}{N}\right) \qquad (2.64)$$

ونسبة التباين في (2.64) إلى (2.62) هي

$$\frac{V(N_j \text{ known})}{V(N_j \text{ not known})} = \frac{S_j^2}{S_j^2 + Q_j \bar{Y}_j^2} = \frac{C_j^2}{C_j^2 + Q_j} \qquad (2.65)$$

حيث $C_j = S_j / \bar{Y}_j$ هو معامل اختلاف الوحدات غير المساوية للصفر. وكما قد نتوقع فإن الانخفاض في التباين العائد لمعرفتنا بـ N_j يكون أكبر عندما تكون نسبة الوحدات المساوية للصفر كبيرة وعندما يكون تغير y_j بين الوحدات غير المساوية للصفر قليلاً نسبيًا. ولمزيد من الدراسة لهذه المسألة انظر Jessen و Houseman (1944) .

(٢-١٤) مقارنات بين متوسطات الميادين

ليكن $\bar{y}_j$ ، $\bar{y}_k$ متوسطي العيّنة في الميدانين j و k من مجموعة الميادين التي صنفنا إليها وحدات عيّنة عشوائية بسيطة. فتباين الفرق بينهما هو

$$V(\bar{y}_j - \bar{y}_k) = V(\bar{y}_j) + V(\bar{y}_k) \qquad (2.66)$$

وتنطبق هذه العلاقة أيضًا على الفرق بين نسبتين $\hat{R}_j$ و $\hat{R}_k$.

وتنبغي ملاحظة نقطة، فمن النادر أن يكون للتساؤل عما إذا كان $\bar{Y}_j = \bar{Y}_k$ أية أهمية علمية، ذلك لأن هذين المتوسطين سوف لا يتساويان بالضبط في مجتمع منته، باستثناء فرصة نادرة الوقوع، وهذا صحيح حتى لو كان البيان الإحصائي في كلي الميدانين مسحوبًا بصورة عشوائية من المجتمع اللانهائي نفسه. وبدلاً من ذلك، نختبر الفرضية الابتدائية بأن الميدانين مسحوبان من مجتمعين لانهائيين لهما المتوسط نفسه. وبالتالي نحذف الـ ت م م عند حساب $V(\bar{y}_j)$ و $V(\bar{y}_k)$ ، مستخدمين العلاقة

$$V(\bar{y}_j - \bar{y}_k) = \frac{S_j^2}{n_j} + \frac{S_k^2}{n_k} \qquad (2.67)$$

وفي حالـة اختبارات الأهمية نحصل على علاقة مشابهة لِـ (2.67) ، إذا صغنا السؤال التالي : هل كان يمكن للعيّنتين من الميدانين أن تُسحبا عشوائيًا من مجتمع منته واحد؟

تحت هذه الفرضية يمكن البرهان [انظر التمرين (٢-١٦)] على أن

$$V(\bar{y}_j - \bar{y}_k) = S_{jk}^2\left(\frac{1}{n_j}+\frac{1}{n_k}\right)$$

حيث S_{jk}^2 تباين المجتمع المنتهي المؤلف من الميدانين معًا.

(٢-١٥) مشروعية التقريب الطبيعي

تأتي الثقة في أن التقريب الطبيعي ملائم لمعظم التطبيقات العملية من مصادر عدة. وقد جرت دراسات كثيرة في نظرية الاحتمال لتوزيع متوسطات عيّنات عشوائية. وقد بُرهن على أنه في أي مجتمع له انحراف معياري منته، ينزع متوسط العيّنة إلى أن يصبح طبيعيًا مع ازدياد n [انظر مثلاً Feller (1957)]. ويتعلق هذا العمل بالمجتمعات اللانهائية.

أمـا في حالـة المعـاينـة بدون إعـادة من مجتمعـات منتهية، فقـد أعـطى Hajek (1960) الشروط الـلازمـة والكـافية التي ينتهي تحتهـا توزيع متوسط عيّنة إلى التـوزيع الـطبيعي، متبعًـا أبحـاث Renyi, Erdös (1959) و Madow (1948) . ويفـترض Hajek متـوالية من القيم n_ν ، N_ν تنتهي إلى الـلانهاية بطريقة ينتهي معها $(N_\nu - n_\nu)$ أيضًا إلى اللانهاية. ونرمز للقياسات في المجتمع ν بِـ $y_{\nu i}$ ، $(i=1,2,\ldots,N)$. وفي هذا المجتمع، لتكن $S_{\nu\tau}$ مجموعة الوحدات في المجتمع التي يكون فيها:

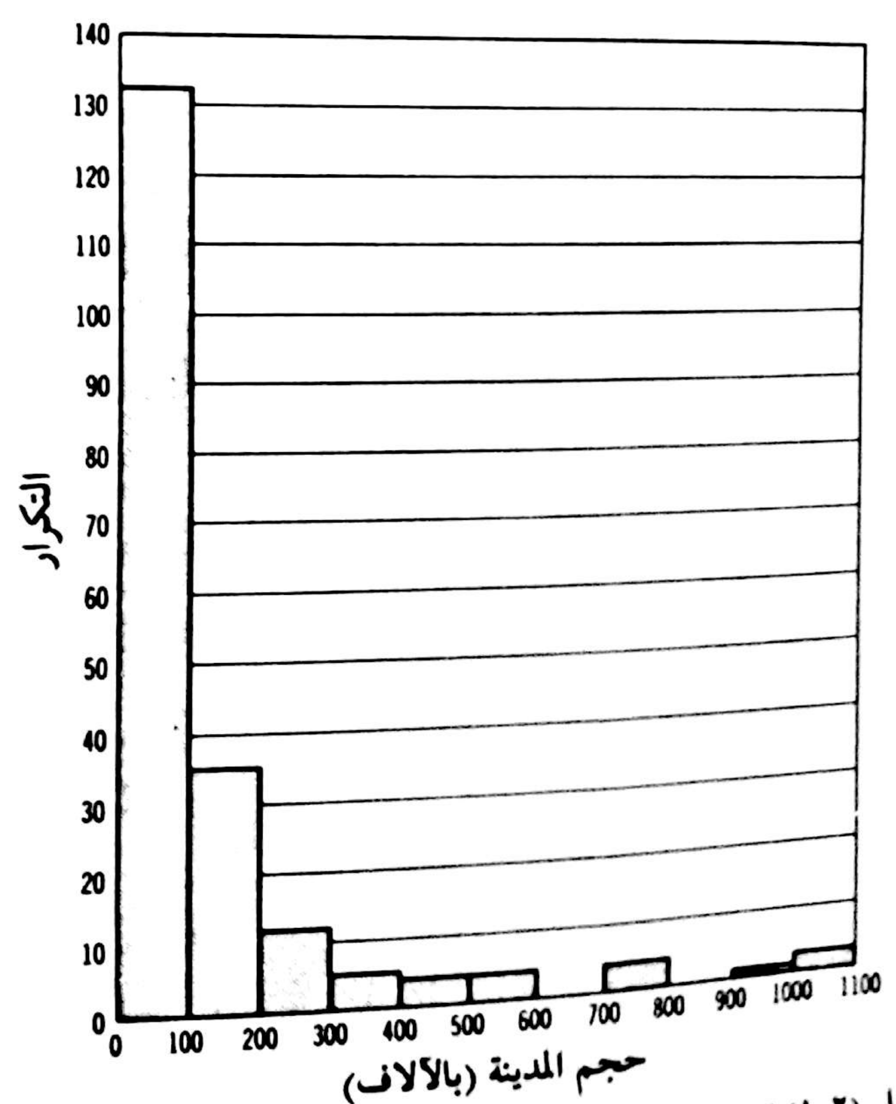
$$|y_{\nu i} - \bar{Y}_\nu| > \tau\sqrt{n_\nu(1-f_\nu)}S_\nu$$

حيث $\bar{Y}_\nu$ ، S_ν ، f_ν هي متـوسط المجتمـع، الانحراف المعياري، والـ ت م م، على الـترتيب، أمـا τ فعـدد موجب. وعنـدئـذ يكـون الشرط من نوع شرط ليـنـدبـرغ (Lindberg) .

$$\lim_{\nu \to \infty} \frac{\sum_{s_{\nu r}} (y_{\nu i} - \bar{Y}_\nu)^2}{(N_\nu - 1)S_\nu^2} = 0$$

لازمًا وكافيًا للتأكيد بأن $\bar{Y}_\nu$ ينتهي إلى التوزيع الطبيعي بالمتوسط والتباين المعطيين في النظريتين (٢ - ١) و (٢ - ٢).

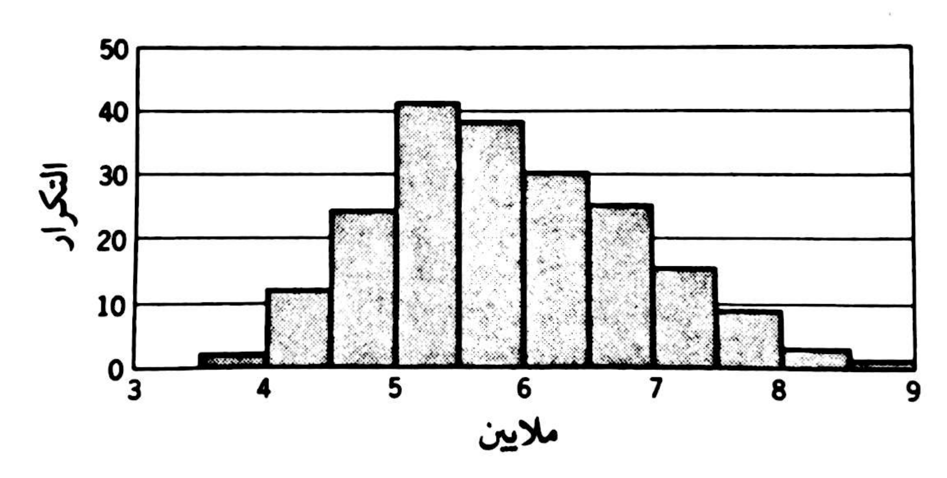
وهذا القدر من المعلومات القيمة يثير الرغبة في أمر ما. فليس من السهل الإجابة عن السؤال المباشر: «في هذا المجتمع، إلى أي حد يجب أن يكون n كبيرًا بحيث يكون التقريب الطبيعي دقيقًا بما فيه الكفاية؟». وتختلف التوزيعات الطبيعية بشدة سواء من حيث طبيعتها أو من حيث درجة بعدها عن التوزيع الطبيعي. وفي التطبيقات العملية للمعاينة، لا يمكن أن نفترض أن توزيعات التكرار التي نواجهها ستكون جميعها قريبة



شكل (٢-١) توزيع التكرار لحجوم 196 من مدن الولايات المتحدة عام 1920 .

بصورة معقولة من التوزيع الطبيعي. فتوزيعات العديد من أنواع المشروعات الاقتصادية (مخازن، مداجن، مدن صغيرة) تُظهر قدرًا كبيرًا من الالتواء، وفي الاتجاه الموجب، أي عدد قليل من الوحدات الكبيرة، والعديد من الوحدات الصغيرة. ويظهر التواء من النوع نفسه في بعض المجتمعات الأحيائية (مثلاً، عدد الفئران أو الذباب على أساس الجادّة الواحدة في مدينة).

وكتوضيح لتوزيع ملتوٍ إيجابًا، يبين الشكل (٢-١) توزيع التكرار لعدد السكان في 196 من المدن الكبرى في الولايات المتحدة عام 1920 . (حُذفت المدن الأربع الكبرى، نيويورك، شيكاغو، فيلاديلفيا، وديترويت، لأن وجودها هنا سيضطرنا إلى تمديد السلم الأفقي إلى أكثر من خمسة أمثال ما هو مبينّ في الشكل، وبالطبع فإنه سيزيد إلى حد كبير من الالتواء). ويبين الشكل (٢-٢) توزيع التكرار للعدد الكلي للسكان في كل من 200 من العيّنات العشوائية البسيطة المسحوبة من هذا المجتمع، حيث $n=49$. وتوزيع مجاميع العيّنات، وكذلك توزيع المتوسطات، أكثر شبهًا بكثير، بالمنحنى الطبيعي، ولكنه لا يزال يُبدي بعضًا من الالتواء وفي الاتجاه الموجب.



شكل (٢-٢) توزيع تكرار مجاميع 200 من العيّنات العشوائية البسيطة حيث n=49 .

ومن نظرية الإحصاء ونتائج تجارب معاينة تناولت مجتمعات ملتوية، يمكن وضع بعض العبارات حول ما يحدث عادةً لاحتمالات الثقة عندما نأخذ عيّنة من مجتمعات ملتوية في اتّجاه موجب، وذلك كما يلي:

١ - تكرار كون العبارة :

$$\bar{y} - 1.96s_{\bar{y}} < \bar{Y} < \bar{y} + 1.96s_{\bar{y}}$$

خاطئة هو عادةً أعلى من 5 بالمائة .

٢ - تكرار كون

$$\bar{Y} > \bar{y} + 1.96s_{\bar{y}}$$

هو أكبر من 2.5 بالمائة .

٣ - تكرار كون

$$\bar{Y} < \bar{y} - 1.96s_{\bar{y}}$$

هو أقل من 2.5 بالمائة .

وكتوضيح ، لنأخذ متغيراً y يتوزع في الأساس وفق التوزيع الثنائي ، بحيث يمكن قراءة التوزيع المضبوط لـ $\bar{y}$ من جداول التوزيع الثنائي . ويأخذ المتغير y قيمتين فقط - القيمة h باحتمال P والقيمة 0 باحتمال Q . ومتوسط المجتمع $\bar{Y}=ph$. وتبينُ عيّنة عشوائية بسيطة حجمها n عدداً من الوحدات الموافقة للقيمة h قدره a و $n-a$ من الوحدات التي تأخذ القيمة صفر. ومن أجل العيّنة

$$\sum y = ah, \qquad \bar{y} = \frac{ah}{n}$$

$$(n-1)s^2 = \sum y^2 - n\bar{y}^2 = ah^2 - \frac{a^2h^2}{n}$$

$$s_{\bar{y}}^2 = \frac{s^2}{n} = \frac{h^2}{n^2}\,\frac{a(n-a)}{n-1}$$

وبالتالي فإننا نقدر 95% حدود ثقة طبيعية لِـ $\bar{Y}$ على الشكل

$$\bar{y} \pm 1.96s_{\bar{y}} = \frac{h}{n}\left[a \pm 1.96\sqrt{\frac{a(n-a)}{n-1}}\right] \qquad (2.68)$$

لتكن $n=400$ ، $p=0.1$ فعندئذ $\bar{Y}=0.1h$. وبالتجربة نجد أنه إذا كان $a=29$ في العبارة (2.68) فحد الثقة الأعلى هو $39.18\ h/400 = 0.098\ h$ بينما يكون $40.34h/400 = 0.01h$ عندما $a=30$. وهكذا فإن أي قيمة لـ a أصغر أو تساوي 29 تعطي حدًّا أعلى للثقة منخفضًا جدًّا. وبصورة مماثلة نجد أنه إذا كان $a\geqslant 54$ فالحد الأدنى يكون مرتفعًا جدًّا.

ويتبع المتغيّر a التوزيع الثنائي حيث $n=400$ ، $P=0.1$. وتبيّن الجداول [Harvard Computation Laboratory, 1955] أن

P_r (الحد الأعلى المعروض منخفض جدًّا) $P_r(a\leqslant 29)=0.0357$

P_r (الحد الأدنى المعروض مرتفع جدًّا) $P_r(a\leqslant 45)=0.0217$

P_r (عبارة ثقة خاطئة) $= 0.0574$

والاحتمال الكلي لكونها خاطئة ليس بعيدًا عن 0.05 وفي أكثر من 60% من العبارات الخاطئة، يكون المتوسط الصحيح أعلى من الحد الأعلى المعروض.

ولا توجد قاعدة عامّة سليمة فيما يتعلق بما يجب أن يكون عليه حجم n لاستخدام التقريب الطبيعي في حساب حدود الثقة. وفي حالة مجتمع يكون انحرافه الرئيس عن التوزيع الطبيعي مؤلفًا من التواء حاد في الاتجاه الموجب وجدتُ من حين لآخر أن القاعدة التالية مفيدة

$$n>25G_1^2 \qquad (2.69)$$

حيث G_1 قياس Fisher للالتواء [Fisher, 1932]

$$G_1=\frac{E(y_i-\bar{Y})^3}{\sigma^3}=\frac{1}{N\sigma^3}\sum_{i=1}^{N}(y_i-\bar{Y})^3 \qquad (2.70)$$

وهذه القاعدة مصمّمة بحيث لا تكون عبارات الـ 95% احتمال ثقة خاطئة في أكثر من 6 بالمائة من المرات. وقد استنبطت رياضيًّا تحت الفرض بإهمال أي اضطراب

يعود إلى عزوم من مرتبة أعلى من 3 لتوزيع $\bar{y}$. وتحاول القاعدة أن تتحكم فقط بمجمل تكرار العبارات الخاطئة، متجاهلة اتجاه خطأ التقدير.

وبحساب G_1 أو، في مجتمع محدد، تقدير لـ G_1 ، نستطيع الحصول على فكرة تقريبية عن حجم العيّنة اللازم لتطبيق التقريب الطبيعي في حساب الثقة. وينبغي التحقق من صحة النتائج، حيثما أمكن ذلك، من خلال تجارب المعاينة.

جدول (٢-٤) توزيع تكرار الفدادين المزروعة في 556 مزرعة

الفئات (بالفدادين)	سلّم رمزي y_i	التكرار f_i	f_iy_i	$f_iy_i^2$	$f_iy_i^3$
0–29	−0.9	47	−42.3	38.1	−34.3
30–63	0	143	0	0	0
64–97	1	154	154	154	154
98–131	2	82	164	328	656
132–165	3	62	186	558	1,674
166–199	4	33	132	528	2,112
200–233	5	13	65	325	1,625
234–267	6	6	36	216	1,296
268–301	7	4	28	196	1,372
302–335	8	6	48	384	3,072
336–369	9	2	18	162	1,458
370–403	10	0	0	0	0
404–437	11	2	22	242	2,662
438–471	12	0	0	0	0
472–505	13	2	26	338	4,394
المجموع		556	836.7	3,469.1	20,440.7

$$E(y_i) = \bar{Y} = \frac{836.7}{556} = 1.50486$$

$$E(y_i^2) = \frac{3469.1}{556} = 6.23939$$

$$E(y_i^3) = \frac{20,440.7}{556} = 36.76385$$

$$\sigma^2 = E(y_i^2) - \bar{Y}^2 = 3.97479$$

$$\kappa_3 = E(y_i - \bar{Y})^3 = E(y_i^3) - 3E(y_i^2)\bar{Y} + 2\bar{Y}^3$$
$$= 15.411$$

$$G_1 = \frac{\kappa_3}{\sigma^3} = \frac{15.411}{7.925} = 1.9$$

مثال

تبيّن المعلومات الإحصائية في الجدول (٢-٤) عدد الفدادين المخصصة للزراعة في 556 مزرعة من منطقة سينيكا - نيويورك. وتأتي هذه المعلومات من سلسلة من الدراسات قام بها West (1951)، الذي سحب عيّنات متكررة حجمها 100 من هذا المجتمع، ودرس دالة تكرار $\bar{y}$ ، s والتوزيع t من أجل عدة متغيّرات لها أهميتها في المسوح الإحصائية المتعلقة بإدارة المزارع.

وحساب G_1 مبيّن تحت الجدول. وقد تمت الحسابات على أساس سلّم قياس رمزي. وبما أن G_1 عدد غير مقيس (ليس له وحدة قياس) فلا حاجة للعودة إلى سلّم القياس الأصلي. ولنلاحظ أن فترة الفئة الأولى تختلف قليلاً عن الفترات الأخرى.

وبما أن $G_1=1.9$ فنأخذ كحدّ أدنى مقترح لـ n العدد:

$$n=(2.5)(1.9)^2=90$$

وقد وجد West من عيّنات حجمها 100 تتعلق بهذا المتغيّر (عدد الفدادين المزروعة) أنه لا توزيع $\bar{y}$ ولا توزيع t يختلف اختلافًا مهمًا عن التوزيعين الطبيعيين النظريين الموافقين.

والممارسات الجيّدة في المعاينة تميل إلى جعل التقريب الطبيعي أكثر صحة. ويفشل التقريب الطبيعي بصورة رئيسة عندما يحوي المجتمع عناصر متطرفة تسيطر في حال وجودها على معدّل العيّنة. وعلى أي حال، فإن للعناصر المتطرفة هذه تأثيراً أكثر خطورة بكثير، وهو زيادة تباين العيّنة وتخفيض الدقة. وبالتالي فإنه من الحكمة عزلها، ووضع خطط منفصلة لمعالجة الموقف، وربما كان ذلك بأخذ تعداد كامل لها عندما لا تكون كثيرة جدًا. وإزاحة هذه العناصر المتطرفة من الجسم الرئيس للمجتمع يخفّض الالتواء ويحسّن التقريب الطبيعي. ومثل هذا التدبير يشكل مثالاً للمعاينة الطبقية التي سنناقشها في الفصل الخامس.

(٢-١٦) المقدَّرات الخطية لمتوسط مجتمع

تحت المعـاينـة العشـوائية البسيطة، هل يكون متوسط العيِّنة $\bar{y}$ أفضل تقدير لِـ $\bar{Y}$ ؟ لقد اجتذب هذا السؤال بصورة طبيعية قدرًا كبيرًا من البحث. ويعتمد الجواب على مجموعة المقدرات المسموح بها والمنافسة لِـ $\bar{y}$ ، وعلى تعريف كلمة «أفضل». ففي المعاينة يجري عادة ترقيم الوحدات في المجتمع بطريقة ما من 1 إلى N . وغالبًا ما تدعى هذه الأرقام التي تميِّز الوحدات عناوين ملحقة بالوحدات.

والنتـائج المبكرة التي برهنها Horvitz و Thompson (1952) في حالة المقدَّرات الخطية في المعاينة العشوائية البسيطة كانت كما يلي: إذا كان أي y_i ينال دائمًا الترجيحة نفسهـا عنـدمـا نسحب وحـدة عنـوانها i ، فمتوسط العيِّنة $\bar{y}$ هو المقدَّر غير المنحاز الوحيد لِـ $\bar{Y}$ من الشكل $\sum^{n} w_i y_i$. وبما أن كل وحدة تظهر في كسر n/N من العيِّنات العـشـوائية البسيطة، فإن $E(\sum^{n} w_i y_i) = n(\sum^{N} w_i\, y_i)/N \equiv \bar{Y}$ فقط إذا كان كل w_i مسـاويًا $1/n$. وبصورة بديلة، إذا كانت الترجيحة تعتمد فقط على الترتيب الذي سحبنا فيه الـوحدة ضمن العيِّنة، فعندئذ يكون لِـ $\bar{y}$ تباين أصغري بين المقدَّرات الخطية غير المنحازة من الشكل $\sum^{n} w_d\, y_{(d)}$ ، حيث $y_{(d)}$ هي قيمة y في الوحدة التي تمخَّض عنها السحب d .

والصف الأوسع من المقدَّرات المنافسة هو المجموعة $\sum^{n} w_{is}\, y_i$ حيث يمكن أن يعتمـد w_{is} على الـوحـدات الأخرى التي تقع ضمن العيِّنة كما يعتمد أيضًا على i . وقـد بيَّن Godambe (1955) أنـه لا يوجـد ضمن هذا الصف تقـدير غير منحـاز لِـ $\bar{Y}$ يتصف بأنه ذو تباين أصغري في جميع المجتمعات.

وقـد اسـتـنـتـج Hartley و J. N. K. Rao (1968, 1969) Royall, (1968) ، و C. R. Rao (1971) ، مزيدًا من الخـواص لِـ $\bar{y}$. وفي أي مجتمع منته سيوجد على الأكثر $T \leqslant N$ من القيم المتميِّزة لِـ y_i . وفي العيِّنة، ليكن هناك n_t من القيم المساوية لِـ y_t حيث $\sum n_t = n$ فيبيِّن Hartley و Rao (1968) أن لِـ $\bar{y}$ تباينًا أصغريًا بين المقدَّرات غير

المنحازة لِـ $\bar{Y}$ التي هي دوالّ في n_i و y_i فقط. ولمعاينة عشوائية مع الإعادة، برهنا أن متوسط القيم المتميزة في العيّنة هو تقدير الإمكانية العظمى لِـ $\bar{Y}$ ، مع أنه ليس ذا تباين أصغري فوق جميع المجتمعات.

وقد درس C. R. Rao (1971) ، متبِّعًا عملاً قام به Kempthorne (1969) ، المقدِّرات غير المنحازة $\hat{\bar{Y}} = \sum w_{is} y_i$ التي درسها Godambe . ولتمثيل الحالة التي لا تقدم فيها العناوين l أية معلومات حول القيم y_l ، قام بحساب معدّل $V(\hat{\bar{Y}})$ فوق جميع التباديل الـ $N!$ للعناوين الملحقة بهذه القيم، وبيّن أن لِـ $\bar{y}$ فوق هذه التباديل تباينًا أصغريًا في المتوسط. أما Royall (1970 b) فقد أعطى نتيجة أعمّ.

كان بحث Godambe (1955) حافزًا للعديد من الأبحاث حول تصميم المعاينة، والتقدير، بما في ذلك موضوعات مثل استخدام معلومات إضافية يمكن أن تحملها العناوين حول القياسات y_i ، ومقدِّرات بايز، وطرق في التقدير نلجأ إليها ونستخدمها عندما يكون ممكنًا وضع افتراضات معينة تتعلق بالتوزيع التكراري لِـ y_i ، وكانت آثار هذه الأعمال على تطبيقات المعاينة محدودة حتى الآن، إلا أنها ينبغي أن تزداد باستمرار. وسنشير إليها من حين لآخر. ولمزيد من المعلومات حول هذا انظر: J. N. K. Rao* (1975 a) و Smith (1976) .

تمارين

(٢-١) في مجتمع يتضمن $N=6$ ، نعلم أن قيم y_i هي 8 ,3 ,1 ,11 ,4 و 7 . احسب متوسط العيّنة $\bar{y}$ في كل من العيّنات العشوائية البسيطة الممكنة التي حجمها 2 . تحقّق أن $\bar{y}$ هو تقدير غير منحاز لِـ $\bar{Y}$ ، وأن تباينه هو كما تعطيه النظرية (٢-٢).

(٢-٢) وللمجتمع نفسه، احسب s^2 لكل من العيّنات العشوائية البسيطة الممكنة التي حجمها 3 وتحقق أن $E(s^2) = S^2$.

---

* من الآن فصاعدًا سيشير لقب Rao في هذا الكتاب إلى J.N.K.Rao مالم نذكر خلاف ذلك.

(٢-٣) إذا كانت العيّنات العشوائية ذات الحجم 2 مسحوبة مع الإعادة من هذا المجتمع، فبيّن عن طريق إيجاد كل العيّنات الممكنة أن $V(\bar{y})$ يحقق المعادلة :

$$V(\bar{y}) = \frac{\sigma^2}{n} = \frac{S^2}{n}\,\frac{(N-1)}{N}$$

(٢-٤) سحبنا عيّنة عشوائية بسيطة من 30 أسرة من مدينة تحوي 14848 أسرة. وكان عدد الأشخاص في كل أسرة من العيّنة كما يلي :

5, 6, 3, 3, 2, 3, 3, 3, 4, 4, 3, 2, 7, 4, 3, 5, 4, 4, 3, 3, 4, 3, 3, 1, 2, 4, 3, 4, 2, 4

قدِّر العـدد الكـلي لسكان المدينة، واحسب احتمال أن يكون هذا التقدير في حدود 10± بالمائة من القيمة الصحيحة.

(٢-٥) في دراسـة لإمكـانية استخدام المعاينة لاختزال العمل في جرد مخزون المستودعات، قمنا بحساب قيمة المواد فوق كل من 36 رفًا في المستودع. وكانت القيم إلى أقرب دولار كما يلي :

29, 38, 42, 44, 45, 47, 51, 53, 53, 54, 56, 56, 56, 58, 58, 59, 60, 60,
60, 60, 61, 61, 61, 62, 64, 65, 65, 67, 67, 68, 69, 71, 74, 77, 82, 85.

وبـاستثـنـاء فرصة واحدة من عشرين (ينبغي للتقدير، الذي نأخذه من عيّنة للقيمـة الإجمالية، أن يكون صحيحًا في حدود 200 $ زيادةً أو نقصانًا. ويرى أحد المستشارين أن عيّنة عشوائية بسيطة من 12 رفًا ستواجه هذه المتطلبات. فهل توافق؟

$$\sum y = 2138, \qquad \sum y^2 = 131{,}682$$

(٢-٦) بعد أخذ العيّنة في الجدول (٢-٢) ص. . . ، قمنا بتعداد الصفحات المملوءة تمامًا (في كل منها 42 توقيعًا) ووجدنا أنها 326 صفحة. استخدم هذه المعلومات لتحسين تقديرك لعدد التواقيع الكلي وأوجد الخطأ المعياري لتقديرك.

(٢-٧) من قائمـة من 468 من الكليات الصغـيرة ذات السنتـين سحبنا عيّنة

عشوائية بسيطة من 100 كلية . وتضمنت العيّنة 54 كلية عامة و 46 كلية خاصة . والبيان المتعلق بعدد الطلاب (y) وعدد المدرسين (x) هو كما يلي :

	n	$\sum(y)$	$\sum(x)$
عـام	54	31,281	2,024
خـاص	46	13,707	1,075

	$\sum(y^2)$	$\sum(yx)$	$\sum(x^2)$
عـام	29,881,219	1,729,349	111,090
خـاص	6,366,785	431,041	33,119

(ا) في كل نوع من الكليات قدِّر نسبة عدد الطلاب إلى عدد الأساتذة .
(ب) احسب الأخطاء المعيارية لتقديراتك .
(جـ) في الكليات العامة ، أوجد 90% حدود ثقة لنسبة الطلّاب إلى المدرسين في المجتمع بكامله .

(٢-٨) في المثال السابق اختبر عند المستوى 5% ما إذا كانت نسبة الطلّاب إلى المدرسين مختلفة بصورة مهمة في النوعين من الكليات .

(٢-٩) في الكليات العامة، قدِّر العدد الكلي للمدرسين (ا) علمًا أن العدد الكلي للكليات العامة في المجتمع هو 251 ، (ب) بدون معرفة هذا الرقم . وفي كل حالة احسب الخطأ المعياري لتقديرك .

(٢-١٠) يبـيـّن الجدول أدناه عدد السكان في كل من 197 من مدن الولايات المتحـدة التي كان عدد سكانها عام 1940 فوق الـ 50,000 . احسب الخطأ المعياري لتقدير عدد السكان الكلي في جميع المدن الـ 197 ، وذلك لكل من الطرق التالية في المعـاينـة : (ا) عيّنة عشوائية بسيطة حجمها 50 ، (ب) عيّنة تحوي المدن الخمس الأكبر و 45 مدينة مختارة على شكل عيّنة عشوائية من بين المدن الـ 192 الباقية ، (جـ) عيّنة تحوي المدن التسع الأكبر و 41 مدينة مختارة كعيّنة عشوائية بسيطة من بين المدن الباقية .

توزيع التكرار لحجوم المدن

f	حجم الفئة f (بالآلاف)	f	حجم الفئة f (بالآلاف)	f	حجم الفئة f (بالآلاف)
....	....	2	550–600	105	50–100
1	1500–1550	1	600–650	36	100–150
....	....	2	650–700	13	150–200
1	1600–1650	0	700–750	6	200–250
....	....	1	750–800	7	250–300
1	1900–1950	1	800–850	8	300–350
....	....	2	850–900	4	350–400
1	3350–3400	0	900–950	1	400–450
....	....	0	950–1000	3	450–500
1	7450–7500	0	1000–1050	0	500–550

ويُشار بِـ . . . للثغرات الموجودة في الفئات.

(٢-١١) احسب معامل الالتواء G_1 للمجتمع الأصلي والمجتمع الباقي بعد إزاحة: (أ) المدن الخمس الأكبر. (ب) المدن التسع الأكبر.

(٢-١٢) سنقوم بمسح إحصائي صغير لمقارنة مالكي المنازل مع المستأجرين. ويوجد في المجتمع حوالي 75% من المالكين و 25% من المستأجرين. ومن أجل أحد مفردات البحث يُعتقد أن التباين هو حوالي 15 لكل من المالكين والمستأجرين. والخطأ المعياري للفرق بين متوسطي الميدانين يجب أن يتجاوز الواحد. كم هو حجم العيّنة التي نحتاجها (أ) إذا أمكن تحديد المالكين والمستأجرين قبل سحب العيّنة، (ب) إذا لم يكن هذا ممكنًا؟ (سيكون الجواب التقريبي مقبولاً في (ب) فالمناقشة الدقيقة تحتاج إلى جداول التوزيع الثنائي).

(٢-١٣) سحبنا عيّنة عشوائية بسيطة حجمها 3 من مجتمع حجمه N مع الإعادة. بيّن أن احتمال أن تتضمن العيّنة 2,1 و 3 من الوحدات المختلفة (مثلاً a,a,a ؛ a,a,b ؛ a,b,c ؛ على الترتيب) هي

$$P_1=\frac{1}{N^2},\qquad P_2=\frac{3(N-1)}{N^2},\qquad P_3=\frac{(N-1)(N-2)}{N^2}$$

وكتقدير لِـ $\bar{Y}$ نأخذ $\bar{y}'$ ، وهو المتوسط غير المرجّح فوق الوحدات المختلفة في العيّنة. بيّن أن معدّل تباين $\bar{y}'$ هو

$$V(\bar{y}')=\frac{(2N-1)(N-1)S^2}{6N^2}\doteq(1-f/2)S^2/3$$

وإحدى الطرق للقيام بذلك هو أن تبيّن أن

$$V(\bar{y}')=S^2\left(\frac{N-1}{N}P_1+\frac{N-2}{2N}P_2+\frac{N-3}{3N}P_3\right)$$

وبالتالي بيّن أن $|V(\bar{y}) < V(\bar{y})$ ، حيث $\bar{y}$ هو المتوسط العادي للملاحظات الـ n في العيّنة. وقد برهن Khamis, DesRaj (1958) أن $V(\bar{y}') < V(\bar{y})$ عندما يكون $n>2$.

(٢-١٤) يقوم طبيبا أسنان A و B بمسح إحصائي يتعلق بحالة أسنان 200 طفل في قرية. ويختار الطبيب A عيّنة عشوائية بسيطة من 20 طفلاً ويُحصي عدد الأسنان المتآكلة عند كل طفل، والنتائج كما يلي:

عدد الأسنان المتآكلة للطفل الواحد	0	1	2	3	4	5	6	7	8	9	10
عدد الأطفال	8	4	2	2	1	1	0	0	0	1	1

ويقوم الطبيب B بالفحص السنّي نفسه بحيث يشمل الأطفال المائتين، مسجلاً فقط أولئك الذين ليس لديهم أسنان متآكلة وقد وجد 60 طفلاً من هذا النوع؛ قدّر العدد الكلي للأسنان المتآكلة عند أطفال القرية، (ا) مستخدمًا نتائج A فقط، (ب) مستخدمًا نتائج A و B ، (جـ) هل التقديرات غير منحازة؟، (د) أي التقديرات تتوقع أن يكون أكثر دقة؟

(٢-١٥) تعتزم شركة مقابلة عيّنة عشوائية بسيطة من المستخدمين الذين أمضوا

في العمل في الشركة أكثر من 5 سنوات. ولدى الشركة 1000 $ لتنفقها، وتكلّف كل مقابلة 10 $. ولا توجد قائمة منفصلة بالمستخدمين الذين عملوا في الشركة لأكثر من 5 سنوات، ولكن يمكن تهيئة قائمة من الملفّات بتكلفة 200 $. ويمكن للشركة إما (ا) إعداد القائمة ومقابلة عيّنة عشوائية بسيطة من المستخدمين الذين يحققون الشرط المطلوب أو (ب) سحب عيّنة عشوائية بسيطة من كل المستخدمين، ثم تقابل من كان منهم محققًا للشرط. ويمكن إهمال تكلفة رفض أولئك الذين لا يحققون الشرط في العيّنة.

بيّن أنه عند تقدير مجموع فوق مجتمع المستخدمين المحققين للشرط، تعطي الخطة (ا) تباينًا أصغر من تباين الخطة (ب) فقط إذا كان $C_j<2\sqrt{Q_j}$، حيث C_j معامل اختلاف المفردة موضع الدراسة بين المستخدمين المحققين للشرط و Q_j نسبة المستخدمين المحققين الشرط في الشركة. تجاهل الـ ت م م.

(٢-١٦) عيّنة عشوائية بسيطة حجمها $n=n_1+n_2$ ومتوسطها $\bar{y}$ سُحبت من مجتمع منته، ثم سُحبت منها عيّنة جزئية عشوائية بسيطة حجمها n_1 ومتوسطها $\bar{y}_1$. بيّن أن (ا) $V(\bar{y}_1-\bar{y}_2)=S^2[(1/n_1)+(1/n_2)]$، حيث $\bar{y}_2$ متوسط الوحدات الـ n_2 الباقية من العيّنة (ب) $V(\bar{y}_1-\bar{y})=S^2[(1/n_1)-(1/n)]$، (ج) $\text{cov}\{\bar{y},\bar{y}_1-\bar{y}\}=0$. وتكرار المعاينة يتضمن إعادة سحب كل من العيّنة والعيّنة الجزئية.

(٢-١٧) عدد العيّنات العشوائية البسيطة المتميزة ذات الحجم n هو بالطبع $N!/n!(N-n)!$. وهناك اهتمام بإيجاد مجموعات أصغر من العيّنات ذات الحجم n والتي تمتلك نفس خواص مجموعة العيّنات العشوائية البسيطة. وإحدى هذه المجموعات هي تلك الخاصة بتصاميم القطاعات غير التامة المتوازنة. وهذه عيّنات تتضمن n وحدة متميزة من N من الوحدات بحيث، (i) تظهر كل وحدة في العدد نفسه (r) من العيّنات، (ii) يظهر كل زوج من الوحدات معًا في λ من العيّنات.

تحقق من أن $\lambda=r(r-1)/(N-1)$ ، وأن عدد العيّنات المتيمزة في المجموعة هو rN/n . وفوق مجموعة عيّنات تصاميم القطاعات غير التامة المتوازنة، برهن بالرموز العادية أنه إذا كان $\bar{y}$ متوسط عيّنة فإن (ا) $V(\bar{y})=(1-f)S^2/n$ ، و(ب) $v(\bar{y})=(1-f)\sum(y_i-\bar{y})^2/n(n-1)$ هو تقدير غير منحاز لـِ $V(\bar{y})$.

ملاحظة : لا توجد طريقة عامة لإيجاد أصغر قيمة لـِ r يمكن معها إقامة تصميم قطاع غير تام متوازن. وأحيانًا يقدم أصغر r معروف $N!/n!(N-n)!$ من العيّنات، مما يعيدنا إلى العيّنات العشوائية البسيطة. ولكن في حالة $N=91$ ، $n=10$ ، يكون لأصغر مجموعة تصاميم قطاعات غير تامة متوازنة 91 عيّنة في مقابل ما يزيد على 6 تريليونات (بـالمقياس الأمـريكي) من العيّنـات العشـوائية البسيطة وقـد بيّن Avadhani و Sukhatme (1973) كيف يمكن استخدام تصاميم القطاعات غير التامة المتوازنة في محاولة لتخفيض تكاليف السفر بين وحدات المعاينة.

(٢ ـ ١٨) ما يلي هو توضيح من Royall (1968) لحقيقة أنه في المعاينة العشوائية البسيطة لا يكون لمتوسط العيّنة $\bar{y}$ تباين أصغري بانتظام ضمن صف المقدّرات من الشكل $\sum^{n} w_{is}y_i$ التي درسها Godambe (1955) ، حيث يمكن للوزن w_i أن يعتمد على الوحدات الأخرى في العيّنة. وفي حالة $N=3$ ، $n=2$ ، خذ المقدّر.

$$\hat{\bar{Y}}_{12}=\tfrac{1}{2}y_1+\tfrac{1}{2}y_2; \qquad \hat{\bar{Y}}_{13}=\tfrac{1}{2}y_1+\tfrac{2}{3}y_3; \qquad \hat{\bar{Y}}_{23}=\tfrac{1}{2}y_2+\tfrac{1}{3}y_3$$

حيث $\bar{Y}_{ij}$ المقـدّر من العيّنـة التي تتضمن الـوحدتين (i,j) . برهن نتائج Royall بأن $\bar{Y}_{ij}$ غير منحاز وأن $V(\bar{Y}_{ij})<V(\bar{y})$ إذا كان $y_3(3y_2-3y_1-y_3)>0$ والتـوضيح مأخـوذ من مثال أسبق لـِ Roy و Chakravarti (1960) .

(٢-١٩) هذا التمرين هو مثال آخر على مقدّرات معدّة من أجل معالم خاصة للمجتمعـات. فعـلى ضوء قرار يجب اتخاذه أخـذت عيّنـة عشـوائية بسيطة، وقـد تبيّن أن y_1 قد يكون منخفضًا بصورة غير عادية و y_N مرتفعًا بصورة غير عادية. وفي حالة

كهذه درس Sarndall (1972) المقدِّر غير المنحاز التالي لِـ $\bar{Y}$.

إذا تضمنت العيّنة y_1دون y_N $\quad \hat{\bar{Y}}_s = \bar{y} + c$

إذا تضمنت العيّنة y_Nدون y_1 $\quad = \bar{y} - c$

في جميع العيّنات الأخرى $\quad = \bar{y}$

حيث c عدد ثابت. برهن نتيجة Sarndall بأن $\hat{\bar{Y}}_s$ غير منحاز وأن

$$V(\hat{\bar{Y}}_s) = (1-f)\left[\frac{S^2}{n} - \frac{2c}{(N-1)}(y_N - y_1 - nc)\right]$$

بحيث يكون $V(\hat{\bar{Y}}_s) < V(\bar{y})$ إذا كان $0 < c < (y_N - y_1)/n$.

(٢-٢٠) عندما يكون $N=8$ في مجتمع و $y_i = 1,4,5,5,\ 6,6,8,13$ ، بيّن أنه في حالة $n=y$ ، $V(\bar{y}) = 1.5$ بينما $V(\hat{\bar{Y}}_s) = 0.214$ عندما $c = 1.5$ (أفضل قيمة لِـ c)، و 0.357عندما تكون c مساوية لِـ 1 و 2 .

وبأخذ المعلومات الواردة في التمرين (٢-١٩) في الاعتبار، يمكن وضع خطة معاينة بديلة فنضع y_1 ، y_8في كل عيّنة، ثم نسحب عيّنة عشوائية بسيطة، حجمها 2 من $y_2, \ldots, y_7$ ، ولنرمز لمتوسط هذه العيّنة بِـ $\bar{y}_2$. وتقدير $\bar{Y}$ هو

$$\hat{\bar{Y}}_{st} = (y_1 + 6\bar{y}_2 + y_8)/8$$

بيّن أن $\hat{\bar{Y}}_{st}$ غير منحاز وأن تباينه $9V(\bar{y}_2)/16$. وفي حالة مجتمع كهذا بيّن أيضًا أن $V(\hat{\bar{Y}}_{st}) = 0.350$. وهذا التقدير مثال لمعاينة طبقية (فصل ٥) بثلاث طبقات هي y_1 ؛ y_2 ، . . . ، y_7و y_8 .

الفصل الثالث

معاينـة النسب والنسب المئويّـة

(٣-١) خواص مميّزة نوعية

نرغب أحيانًا في تقدير العدد الكلي، أو النسبة المئوية لوحدات في المجتمع تمتلك خاصة مميّزة ما، أو انتماء، أو تقع ضمن صف معرَّف. والعديد من النتائج التي تُنشر بانتظام من عمليات حصر شامل، أو مسوح إحصائية، هي من هذا القبيل، فمثلاً عدد الأشخاص العاطلين عن العمل، النسبة المئوية من المجتمع التي تتصف بأنها وطنية المولد. ويمكن إدخال التصنيف مباشرة إلى ورقة الاستبيان، كما في المسائل التي يُجاب عنها بـ«نعم» أو «لا». وفي حالات أخرى تكون القياسات الأصلية مستمرة بشكل أو بآخر، ويجري إدخال التصنيف عند جدولة النتائج. فيمكن أن نسجل مثلاً أعمار المستجيبين إلى أقرب سنة، ولكننا ننشر النسبة المئوية في المجتمع لمن هم في سن الستين وما فوق.

رموز: نفرض أن كل وحدة في المجتمع تقع في أحد الصفّين C أو C'. والرمز كما يلي:

عدد الوحدات من C في		نسبة الوحدات من C في	
العيّنة	المجتمع	العيّنة	المجتمع
$p=a/n$	$P=A/N$	a	A

وتقدير العيّنة لِـ P هو p، وتقدير العيّنة لِـ A هو Np أو Na/n. وفي العمل الإحصائي، غالبًا ما يجري تطبيق التوزيع الثنائي على تقديرات مثل a و p. وكما سنرى فإن التوزيع

الصحيح في حالة مجتمعات منتهية هو التوزيع فوق الهندسي، ومع ذلك فإن التوزيع الثنائي يشكّل عادةً تقريبًا مُرضيًا.

(٣-٢) تباين تقديرات العيّنة

يمكننا، من خلال تدبير بسيط، تطبيق النظريات التي أرسيناها في الفصل الثاني على هذه الحالة. ففي كل وحدة من العيّنة أو المجتمع، لنعرِّف y_i على أنه 1 إذا كانت الوحدة من C وعلى أنه صفر إذا كانت من C'. ومن الواضح أنه في هذا المجتمع من قيم y_i:

$$Y = \sum_1^N y_i = A \qquad (3.1)$$

$$\bar{Y} = \frac{\sum_1^N y_i}{N} = \frac{A}{N} = P \qquad (3.2)$$

ولدينا أيضًا في العيّنة:

$$\bar{y} = \frac{\sum_1^n y_i}{n} = \frac{a}{n} = p \qquad (3.3)$$

وبالتالي يمكن اعتبار مسألة تقدير A و P كتلك المتعلقة بتقدير مجموع ومتوسط مجتمع يأخذ فيه كل y_i القيمة 1 أو صفرًا. ولكي نستخدم نظريات الفصل الثاني، نعبّر أولاً عن S^2 و s^2 بدلالة P و p. ونلاحظ أن،

$$\sum_1^N {y_i}^2 = A = NP, \qquad \sum_1^n {y_i}^2 = a = np$$

ومنه:

$$S^2 = \frac{\sum_1^N (y_i - \bar{Y})^2}{N-1} = \frac{\sum_1^N {y_i}^2 - N\bar{Y}^2}{N-1}$$

$$= \frac{1}{N-1}(NP - NP^2) = \frac{N}{N-1}PQ \qquad (3.4)$$

حيث $Q=1-p$. وبصورة مماثلة :

$$s^2 = \frac{\sum_{1}^{n} (y_i - \bar{y})^2}{n-1} = \frac{n}{n-1} pq \qquad (3.5)$$

ويعطي تطبيق النظريات (٢-١)، (٢-٢) و(٢-٤) على هذا المجتمع ، النتائج التالية في حالة معاينة عشوائية بسيطة للوحدات التي تمَّ تصنيفها .

نظرية (٣-١)

نسبة العيّنة $p=a/n$ هي تقدير غير منحاز لنسبة المجتمع $P=A/N$.

نظرية (٣-٢)

تباين p هو

$$V(p) = E(p-P)^2 = \frac{S^2}{n}\left(\frac{N-n}{N}\right) = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right) \qquad (3.6)$$

وذلك باستخدام المعادلة (3.4) .

نتيجة (١)

إذا كانت p و P النسبتين المئويتين من العيّنة والمجتمع ، على الترتيب ، للوحدات الواقعة في الصف C ، فتستمر صحة العلاقة (3.6) من أجل تباين p .

نتيجة (٢)

تباين $\hat{A}=Np$. وهو المجموع المقدَّر لوحدات الصف C ، يساوي

$$V(\hat{A}) = \frac{N^2 PQ}{n}\left(\frac{N-n}{N-1}\right) \qquad (3.7)$$

نظرية (٣-٣)
التقدير غير المنحاز لتباين p كما نحسبه من العيّنة هو

$$v(p) = s_p{}^2 = \frac{N-n}{(n-1)N} pq \tag{3.8}$$

برهان
بيّنا في النتيجة التالية للنظرية (٢-٤) أنه في حالة متغير مستمر y_i يكون :

$$v(\bar{y}) = \frac{s^2}{n} \frac{(N-n)}{N} \tag{3.9}$$

تقديرًا غير منحاز لتباين متوسط العينة $\bar{y}$.

وفي حالة النسب، تأخذ p مكان $\bar{y}$ ، وقد بيّنا في المعادلة (3.5) أن

$$s^2 = \frac{n}{n-1} pq \tag{3.10}$$

ومنه :

$$v(p) = s_p{}^2 = \frac{N-n}{(n-1)N} pq \tag{3.11}$$

وينتج أنه إذا كان N كبيرًا جدًّا بالنسبة إلى n ، بحيث يمكن إهمال الـ(ت م م).
فإن الكمية

$$\frac{pq}{n-1}$$

هي تقدير غير منحاز لتباين p .
وقد تبدو هذه النتيجة محيّرة لبعض القرّاء، باعتبار أن العبارة pq/n مستخدمة بصورة ثابتة تقريبًا، في التطبيقات العملية، من أجل تقدير التباين. والحقيقة هي أن pq/n ليس تقديرًا غير منحاز حتى في حالات مجتمعات لا نهائية.

نتيجة

المقدار

$$v(\hat{A}) = s_{Np}{}^2 = \frac{N(N-n)}{n-1} pq \qquad (3.12)$$

هو تقدير غير منحاز لتباين $\hat{A}=Np$ ، وهو تقدير لمجموع عدد وحدات المجتمع الواقعة في الصف C .

مثال

من قائمة تحوي 3042 من الأسماء والعناوين، بيّنت عيّنة عشوائية بسيطة من 200 من الأسماء، لدى التقصّي، أن 38 من العناوين خاطئة. قدِّر العدد الكلي للعناوين التي تحتاج إلى تصحيح في القائمة، واحسب الخطأ المعياري لهذا التقدير.
لدينا:

$$N=3042, \qquad n=200, \qquad a=38, \qquad p=0.19$$

وتقدير العدد الكلي للعناوين الخاطئة هو

$$\hat{A} = Np = (3042)(0.19) = 578$$

$$S_{\hat{A}} = \sqrt{[(3042)(2842)(0.19)(0.81)/199]} = \sqrt{6686} = 81.8$$

وبما أن نسبة المعاينة تحت السبعة في المائة، فإن الأثر الذي يتركه عامل الـ(ت م م) بسيط جدًّا، ولإزاحة هذا العامل نضع N بدلاً من $(N-n)$. وبالإضافة إلى ذلك، إذا وضعنا n بدلاً من $(n-1)$ نجد العلاقة الأبسط

$$s_{Np} = N\sqrt{pq/n} = (3042)\sqrt{(0.19)(0.81)/200} = 84.4$$

وهذه النتيجة على توافق قريب بصورة مقبولة من النتيجة السابقة وهي 81.8 .

وتصح العلاقات السابقة الموافقة لتباين وتقدير تباين p ، فقط إذا صُنِّفت الوحدات إلى C أو C' وبحيث تكون p هي نسبة عدد وحدات العيّنة الواقعة في C إلى العدد الكلي لوحدات العيّنة. وتوجد حالة عامة تتألف فيها كل وحدة من مجموعة من

العناصر، والعناصر هي التي يجري تصنيفها. وفيما يلي قليل من الأمثلة :

وحدة المعاينة	العناصر
أسرة	أعضاء الأسرة
مطعم	مستخدمون
صندوق بيض	البيضات منفردة
شجرة خوخ	ثمرات الخوخ منفردة

إذا سحبنا عيّنة عشوائية بسيطة من الوحدات بهدف تقدير النسبة P لعناصر المجتمع التي تنتمي إلى الصف C ، فلا يمكن تطبيق العلاقات السابقة، ونقدم طُرقًا مناسبة في فقرة لاحقة.

(٣-٣) تأثير P على الأخطاء المعيارية

تبين المعادلة (3.6) كيفية تغير تباين تقدير نسبة مع تغيّر P، وذلك من أجل قيم ثابتة لـ n و N . وإذا أهملنا عامل الـ(ت م م) نجد

$$V(p) = \frac{PQ}{n}$$

ويبين الجدول (٣-١) الدالة PQ وجذرها التربيعي. ويمكن اعتبار هاتين الدالتين تباين عيّنة حجمها الواحد، وانحرافها المعياري، على الترتيب.

جدول (٣-١) قيم PQ و $\sqrt{PQ}$ حيث P هي النسبة المئوية من المجتمع الواقعة في الصف C

P	0	10	20	30	40	50	60	70	80	90	100
PQ	0	900	1600	2100	2400	2500	2400	2100	1600	900	0
$\sqrt{PQ}$	0	30	40	46	49	50	49	46	40	30	0

وتأخذ الدالتان قيمتهما العظمى عندما ينقسم المجتمع على التساوي بين الصفّين، وهما متناظران حول هذه النقطة. ويتغير الخطأ المعياري لـ p بصورة طفيفة

نسبيًا عندما تكون p في أي مكان بين 30 بالمائة و 70 بالمائة. ونحتاج عند القيمة العظمى لِـ $\sqrt{PQ}$ وهي 50 ، إلى عيّنة حجمها 100 ، لكي ينخفض الخطأ المعياري للتقدير إلى 5 بالمائة. وبلوغ خطأ معياري مساوٍ لِـ 1 بالمائة يحتاج إلى عيّنة حجمها 2500 .

ولا تكون هذه المعالجة مناسبة عندما ينصبّ اهتمامنا على العدد الكلي لوحدات المجتمع الواقعة ضمن الصف C . ومن الطبيعي أكثر، في هذه الحالة، أن نسأل: هل يمكن أن يكون التقدير صحيحًا في حدود 7 بالمائة من المجموع الحقيقي ، مثلاً؟ وهكذا ننحو إلى التفكير في الانحراف المعياري معبرًا عنه على شكل كسر أو نسبة مئوية من القيمة الحقيقية NP والكسر هو

$$\frac{\sigma_{Np}}{NP}=\frac{N\sqrt{PQ}}{\sqrt{n}NP}\sqrt{\frac{N-n}{N-1}}=\frac{1}{\sqrt{n}}\sqrt{\frac{Q}{P}}\sqrt{\frac{N-n}{N-1}} \qquad (3.13)$$

وتدعى هذه الكمية عادةً معامل اختلاف التقدير. وإذا كان عامل (ت م م) مهملاً يصبح هذا المعامل مساويًا لِـ $\sqrt{Q/nP}$. ويبين الجدول (٣-٢) النسبة $\sqrt{Q/P}$ التي يمكن اعتبارها معامل الاختلاف لعيّنة حجمها 1 .

جدول (٣-٢) قيم $\sqrt{Q/P}$ عند قيم مختلفة لِـ P حيث P هي نسبة المجتمع المئوية الواقعة في الصف C

P	0	0.1	0.5	1	5	10	20
$\sqrt{Q/P}$	∞	31.6	14.1	9.9	4.4	3.0	2.0
P	30	40	50	60	70	80	90
$\sqrt{Q/P}$	1.5	1.2	1.0	0.8	0.7	0.5	0.3

وفي حالة عيّنة حجمها ثابت، يتناقص معامل اختلاف تقدير المجموع الذي يقع ضمن الصف C بصورة ثابتة كلما ازدادت النسبة المئوية الحقيقية ضمن C . ويكون

المعامل مرتفعًا عندما تكون P أقل من 5 بالمائة. ونحتاج لعيّنات كبيرة جدًا كي نحصل على تقديرات دقيقة للعدد الكلي الذي يمتلك أي صفة من الصفات النادرة في المجتمع. وفي حالة $P=1\%$ يجب أن يكون $\sqrt{n}=99$ حتى ينخفض معامل اختلاف التقدير إلى 0.10 أو 10 بالمائة. وهذا يعني أن حجم العيّنة يساوي 9801. وتصبح طريقة المعاينة العشوائية البسيطة، أو أية طريقة في المعاينة معدّة لأغراض عامة، طريقة كثيرة التكاليف عند تقدير العدد الكلي للوحدات من النوع النادر.

(٣-٤) التوزيع الثنائي

بما أن المجتمع هو من نوع خاص مبسّط، تأخذ فيه y_i إما القيمة 1 أو 0 فيمكننا إيجاد دالة التكرار الفعلية للتقدير p وليس فقط متوسطه وتباينه.

ويحوي المجتمع A من وحدات الصف C و $(N\text{-}A)$ وحدة من C' ، حيث $P=A/N$. وإذا اتفق أن كانت الوحدة المسحوبة الأولى من الصف C ، فيبقى في المجتمع $(A-1)$ وحدة من C و $(N-A)$ وحدة من C' . وهكذا تتغير نسبة الوحدات من C بعد السحب الأول، بصورة طفيفة فتصبح $(A\text{-}1)/(N\text{-}1)$. وعلى الوجه الآخر، إذا كان السحب الأول من C' تتحول نسبة الوحدات من C إلى $A/(N\text{-}1)$. وفي المعاينة بدون إعادة تبقى النسبة في تغير مستمر حتى ينتهي السحب بأكمله. وفي هذه الفقرة نتجاهل مثل هذه التغيّرات، أي أننا نفرض P ثابتة، وهذا يؤدي إلى الفرض بأن كلًا من A و $(N\text{-}A)$ كبير بالنسبة لحجم العيّنة n ، أو أن المعاينة هي معاينة مع الإعادة.

وتحت هذا الفرض، تتألف عملية سحب العيّنة من سلسلة من n من التكرارات، واحتمال سحب وحدة من C يساوي في كل منها P . وهذه الحالة تُنتج دالة التكرار الثنائية المعروفة لعدد وحدات العيّنة التي تنتمي إلى الصف C . واحتمال أن تحوي العيّنة a من وحدات الصف C هو

$$\Pr(a)=\frac{n!}{a!(n-a)!}P^{a}Q^{n-a} \tag{3.14}$$

ويمكن جدولة دالة تكرار a أو $p=a/n$ ، أو تقدير المجموع NP مستخدمين هذه العبارة.

هنـاك ثلاث مجمـوعـات شاملة من الجداول. وجميعها تعرض P وفق فترات طولها 0.01 . ومدى n هو كما يلي في كل منها.

مكتب الولايات المتحدة للمقاييس U.S. Bureau of Satandards (1950) .

(أي أنها تذهب من 1 إلى 49 وفق فترات تساوي الواحد) $N=1(1)49$

Roming (1952) :

$n=50(5)100$

معمل هارفارد للحسابات Havard Computation Laboratory (1955) :

$n=1\ (1)50\ (2)100\ (10)200\ (20)500\ (50)1000$

(٣-٥) التوزيع فوق الهندسي

يمكن إيجاد توزيع p دون الفرض بأن المجتمع كبير بالنسبة إلى العيّنة. إن عدد وحدات المجتمع في الصفين C و C' هو A و A' ، على الترتيب. وسنحسب احتمال أن الأعداد الموافقة في العيّنة هي a و a' ، على الترتيب، حيث

$$A+A'=N \qquad \text{و} \quad a+a'=n$$

وفي المعـاينـة العشـوائية البسيطة يكـون لكـل من الـ $\binom{N}{n}$ من الاختيارات المختلفة لـ n وحدة من أصل N الفرصة نفسها في أن تكون هي الوحدات المسحوبة. ولإيجـاد الاحتـمال الـذي نريده، نحسب كم من هذه العيّنـات يتضمن a وحـدة بالـضـبط من C و a' من C' . وعـدد الاختيارات المختلفـة لـ a من الـوحـدات الـ A المـوجـودة في C هو $\binom{A}{a}$ بينـما عدد الاختيارات المختلفـة لـ a' من أصـل A' هو $\binom{A'}{a'}$. ويمكن ضم أي اختيار من النوع الأول لأي اختيار من النوع الثاني لتشكيل عيّنة مختلفة بالصورة المطلوبة. ويكون العدد الكلي للعيّنات من الشكل المطلوب هو إذن

$$\binom{A}{a}\cdot\binom{A'}{a'}$$

ومنه إذا سحبنا عيّنة عشوائية بسيطة حجمها n ، فإن احتمال كونها من النوع المطلوب هو

$$Pr(a, a'|A, A') = \binom{A}{a} \cdot \binom{A'}{a'} \Big/ \binom{N}{n} \tag{3.15}$$

وهـذا هو التـوزيع التكراري لِـ a أو np ، ومنه مباشرة يمكن استنتاج توزيع p ويدعى التوزيع في (3.15) *بالتوزيع فوق الهندسي*.

ولتسهيل الحسـابـات يمكن كتـابـة الاحتمالات فوق الهندسية في (3.15) على الشكل التالي

$$\frac{n!}{a!(n-a)!} \cdot \frac{A(A-1)\ldots(A-a+1)(A')(A'-1)\ldots(A'-a'+1)}{N(N-1)\ldots(N-n+1)} \tag{3.16}$$

مثال

أسرة من 8 تحوي 3 ذكـور و 5 إنـاث. أوجد دالة تكرار عدد الذكور في عيّنة عشوائية بسيطة حجمها 4 . في هذه الحالة لدينا

$A=3$ و $A'=5$ و $N=5$ ، $n=4$ ،

ومن العلاقة (3.16) نجد أن دالة تكرار عدد الذكور a هو كما يلي:

a	الاحتمال
0	$\frac{4!}{0!4!} \cdot \frac{5.4.3.2}{8.7.6.5} = \frac{1}{14}$
1	$\frac{4!}{1!3!} \cdot \frac{3.5.4.3}{8.7.6.5} = \frac{6}{14}$
2	$\frac{4!}{2!2!} \cdot \frac{3.2.5.4}{8.7.6.5} = \frac{6}{14}$
3	$\frac{4!}{3!1!} \cdot \frac{3.2.1.5}{8.7.6.5} = \frac{1}{14}$
4	مستحيلة = 0

ويمكن أن يتحقق القارىء من أن متوسط عدد الذكور هو 3/2 وتباينه 15/28 . وتتفق هذه النتائج مع القانونين المذكورين سابقًا، في الفقرة (٣-٢)، واللذان يعطيان

$$E(np) = nP = \frac{nA}{N} = \frac{(4)(3)}{8} = \frac{3}{2}$$

$$V(np) = nPQ\frac{N-n}{N-1} = 4 \cdot \frac{3}{8} \cdot \frac{5}{8} \cdot \frac{4}{7} = \frac{15}{28}$$

(٣-٦) حدود الثقـة

نناقش أولاً معنى حدود الثقـة في حالـة مميزات وصفية. في العيّنـة يقـع a من أصل n في الصف C . لنفرض أن الاستقراءات ستتم حول العدد A من وحدات المجتمـع الـواقعـة ضمن الصف C وللحصـول على حد الثقـة الأعلى لـ A نحسب قيمة $\hat{A}_U$ بحيث يكون احتمال أن تحوي العيّنة a أو أقل من وحدات C ، من أجل هذه القيمة عبارة عن مقدار صغير α_U مثلاً 0.025 . وبصورة أدق تحقق $\hat{A}_U$ المعادلة.

$$\sum_{j=0}^{a} \Pr(j, n-j | \hat{A}_U, N-\hat{A}_U) = \alpha_U U \qquad (3.17)$$

حيث Pr هو الحـد الاحتـمالي المـوافق للتـوزيع فوق الهنـدسي، كما عرفنـاه في المعادلة (3.15) .

عنـدمـا نختـار α_U سلفًـا، يتطلب تحقق المعادلة (3.17) ، قيمة غير صحيحة لِـ $\hat{A}_U$ ، بصورة عامة، بينما ينبغي أن تكون A عددًا صحيحًا. وعمليًا نختار A_U مساوية لأصغـر قيمـة صحيحـة لِـ A بحيث يكـون الـطرف الأيسر من (3.17) أقـل من أو يساوي α_U. وبصورة مشابهة، فإن حد الثقة الأدنى $\hat{A}_L$ هو أكبر قيمة صحيحة بحيث إن

$$\sum_{j=a}^{n} \Pr(j, n-j | \hat{A}_L, N-\hat{A}_L) \le \alpha_L \qquad (3.18)$$

وعندئذ نجد حدود الثقة لِـ P بأخذ $\hat{P}_L = \hat{A}_L/N$. و $\hat{P}_U = \hat{A}_U/N$,

وتتوافر طرق كثيرة جدًا لحساب حدود الثقة.

طرق مضبوطة تمامًا

يقدم Chung و Delury (1950) جداول لِـ 90, 95 و 99% حدود ثقة لِـ P ، وذلك في حالة N=500 ، N=2500 و N=10000 ويمكن الحصول على القيم الموافقة لمجتمعات

من حجوم فيما بين هذه الحجوم بوساطة الاستيفاء. ويعطي Lieberman و Owen (1961) جداول تتضمن حدودًا فردية وتراكمية للتوزيع فوق الهندسي، إلا أن N يمتد فقط حتى المائة.

التقريب الطبيعي

من (3.8) الخاصة بتقدير تباين p، نجد أن أحد أشكال التقريب الطبيعي لحدود ثقة لِـ P هو

$$p \pm \left[t\sqrt{1-f}\sqrt{pq/(n-1)} + \frac{1}{2n} \right] \qquad (3.19)$$

حيث $f=n/N$ و t قيمة المتغير الطبيعي الموافقة لاحتمال الثقة. ونادرًا ما يقدم استخدام الحد المألوف $u\sqrt{pq / n}$ فرقًا يُذكر. والحد الأخير في الطرف الأيمن هو تصحيح من أجل الاستمرار. ويُنتج هذا تحسنًا طفيفًا فقط في دقة التقريب.

وعلى أية حال، يعطي التقريب الطبيعي عادةً، وبدون التصحيح، فترة ثقة ضيقة تمامًا. ويتوقف الخطأ في التقريب الطبيعي على المقادير N,p,n وα_U وα_L. والكمية التي يكون الخطأ فيها أكثر حساسية هي np، أو، على وجه التحديد، العدد الملحوظ من الصف الأصغر. ويقدم الجدول (٣-٣) قواعد عمل لتقرير متى يمكن استخدام التقريب الطبيعي (3.19).

جدول (٣-٣) أصغر قيم لِـ np تبرّر استخدام التقريب الطبيعي

n = حجم العيّنة	np = عدد الملاحظات في الصف الأصغر	p
30	15	0.5
50	20	0.4
80	24	0.3
200	40	0.2
600	60	0.1
1400	70	0.05
∞	80	~0*

* هذا يعني أن p صغيرة جدًا بحيث يتبع np توزيع بواسون

وقد وُضعت القواعد في الجدول (٣-٣) بحيث إنه مع 95% حدود ثقة، لا تزيد القيمة الحقيقية لاحتمال أن يفشل حدا الثقة في احتواء القيمة p عن 5.5% ، بالإضافة إلى أن احتمال أن يكون الحد الأعلى أقل من P هو بين 2.5 و 3.5 بالمائة، واحتمال أن يتجاوز الحد الأدنى P يقع بين 2.5 و 1.5 بالمائة.

مثال (١)

يوجد 37 من وحدات الصف C في عيّنة عشوائية بسيطة حجمها 100 مسحوبة من مجتمع حجمه 500 . أوجد 95% حدي ثقة لنسبة ومجموع وحدات الصف الموجودة في المجتمع. لدينا في هذا المثال

$$p=0.37 \qquad N=500 \qquad n=100$$

ويقع المثال ضمن الحدود التي تسمح باستخدام التقريب الطبيعي. والخطأ المعياري المقدر لـ p هو

$$\sqrt{(1-f)pq/(n-1)}=\sqrt{(0.8)(0.37)(0.63)/99}=0.0434$$

والتصحيح من أجل الاستمرار وهو $1/2n$ يساوي 0.005 . ولذلك فإن تقدير حدي الثقة لـ P هو

$$0.37\pm(1.96\times0.0434+0.005)=0.37\pm0.090$$

$$\hat{P}_L=0.280, \qquad \hat{P}_U=0.460$$

والحدان كما نقرؤهما من جداول Chung و Delury هما 0.285 و 0.462 على الترتيب.

ولإيجاد الحدود الخاصة بالعدد الكلي لوحدات الصف C في المجتمع، نضرب بـ N فنحصل على 140 و 230 على الترتيب.

تقريبات التوزيع الثنائي

عندما لا ينطبق التقريب الطبيعي، يمكن إيجاد حدود لـ P من جداول التوزيع الثنائي (فقرة ٣-٤) وتعديلها عند الضرورة لتأخذ في الحساب عامل الـ ت م م. ويعطي الجدول VIII_1 في الجداول الإحصائية لـ Fisher و Yates (1957) حدود

ثقة لِـ P باستخدام التوزيع الثنائي، وذلك من أجل أية قيمة لِـ n ، وهي بديل مفيد للجداول العادية للتوزيع الثنائي. ويبين المثال (٢) كيفية حساب التقريب الثنائي.

مثال (٢)

من أجل مفردة أخرى في العيّنة الواردة في المثال (١)، يقع 9 من أصل 100 وحدة في الصف C . ومن جدول Romig في حالة $n=100$ ، نجد أن 95% حدي ثقة لِـ P هما 0.041 و 0.165 . (تعطي جداول Fisher - Yates 0.042 و 0.164) وإذا كان f كسر المعاينة أقل من 5% ، تكون الحدود المحسوبة بهذه الطريقة دقيقة بما فيه الكفاية من أجل معظم الأهداف. وفي هذا المثال $f=0.2$ ونحتاج إلى التعديل.
ولإجراء التعديل، نقصر الفترة بين p وكل من الحدين بنسبة $\sqrt{1-f}=\sqrt{0.8}=0.894$.
وتكون الحدود المعدلة كما يلي:

$$\hat{P}_L = 0.090-(0.894)(0.090-0.041)=0.046$$
$$\hat{P}_U = 0.090+(0.894)(0.165-0.090)=0.157$$

والحدان كما نقرؤهما من جداول Chung و Delury هما 0.045 و 0.157 على الترتيب.

وقدم Burstein (1975) بديلاً لهذه الحسابات أكثر دقة بقليل. فلنفرض أن a وحدة من أصل n من الوحدات تقع في الصف C (في هذا المثال، $a=9$ ، $n=100$). في $\hat{P}_L$ نضع $(a-0.5)/n=0.085$ بدلاً من $a/n=0.090$. وفي $\hat{P}_U$ ، نضع $(a+a/n)=0.0909$ ونأخذ أيضًا $(1-f)$ على الشكل $(N-n)/(N-1)$ وهكذا نجد وفقًا لطريقة Burstein أن

$$\hat{P}_L = 0.085-(0.895)(0.085-0.041)=0.046$$
$$\hat{P}_U = 0.0909+(0.895)(0.165-0.0909)=0.157$$

ولا توجد تغيرات في الحدين، في هذا المثال.

مثال (٣)

وفي التسجيلات السمعية التي نطلب فيها معدّل خطأ منخفضًا جدًا، نهتم في

المقام الأول بحد الثقة الأعلى لِـ A . لنفرض أن 200 تسجيل من أصل 1000 قد تم التحقق منها وأننا نقبل التسجيلات الألف إذا لم نعثر على أيّة أخطاء. فهناك جداول موضوعة كي تعطي حد الثقة الأعلى لعدد الأخطاء في الدفعة من التسجيلات بأكملها. ونستنتج تقريبًا جيدًا من العلاقة التالية. احتمال عدم العثور على خطأ في n ـ علماً بأن A من الأخطاء موجودة في N ـ هو وفقًا للتوزيع فوق الهندسي.

$$\frac{(N-A)(N-A-1)\dots(N-A-n+1)}{N(N-1)\dots(N-n+1)} \doteq \left(\frac{N-A-u}{N-u}\right)^n$$

حيث $u=(n-1)/2$. وعلى سبيل المثال، في حالة $n=200$ ، $A=10$ ، $N=1000$ ، يعطي التقريب $(890.5/900.5)^{200}$ ، وباللوغاريتمات نجد أنها تساوي 0.107 وهكذا يكون $A=10$ (معدل خطأ يساوي 1%) هو على وجه التقريب 90% حد ثقة أعلى لعدد الأخطاء في دفعة التسجيلات الألف.

(٣-٧) التصنيف في أكثر من صفين

عند تقديم النتائج، كثيراً ما نصنف الوحدات إلى أكثر من صفين. إذ يمكن تصنيف عيّنة من مجتمع بشري إلى خمس عشرة زمرة مدى كل منها خمس سنوات. وحتى عندما يكون الجواب المفترض لسؤال هو «نعم» أو «لا» فيمكن أن تقع النتائج التي نحصل عليها فعلاً في أربعة صفوف «نعم»، «لا»، «لا أعرف»، « لا جواب». وسنوضح تعميم النظرية إلى مثل هذه الحالات، من خلال الحالة التي يوجد فيها ثلاثة صفوف. ونفترض أن عدد الوحدات، من الصف i ، هو A_i في المجتمع و a_i في العيّنة. حيث

$$N=\sum A_i, \qquad n=\sum a_i, \qquad P_i=\frac{A_i}{N}, \qquad p_i=\frac{a_i}{n}$$

وعندما يكون حجم العيّنة صغيراً بالنسبة لكل المقادير A_i ، فيمكن، بصورة ناجعة، اعتبار الاحتمالات P_i ثابتة خلال عملية سحب العيّنة واحتمال سحب العيّنة الملحوظة فعلاً معطى من خلال عبارة التوزيع متعدد الحدود.

$$\Pr(a_i)=\frac{n!}{a_1!a_2!a_3!}P_1^{a_1}P_2^{a_2}P_3^{a_3} \qquad (3.20)$$

وهو التعميم المناسب للتوزيع الثنائي، ويشكّل تقريبًا جيدًا عندما يكون كسر المعاينة صغيراً.

والعبارة الصحيحة لاحتمال سحب العيّنة الملحوظة هي

$$\Pr(a_i|A_i) = \binom{A_1}{a_1}\binom{A_2}{a_2}\binom{A_3}{a_3}\Big/\binom{N}{n} \qquad (3.21)$$

وهذه العبارة هي الامتداد الطبيعي للمعادلة (3.15) فقرة (٣-٥)، والموافقة للتوزيع فوق الهندسي. فالبسط هو عدد العيّنات المتميزة ذات الحجم n التي يمكن تشكيلها متضمنة a_i وحدة من الصف 1، a_2 وحدة من الصف 2 و a_3 وحدة من الصف 3 .

(٣ - ٨) حدود الثقة عند وجود أكثر من صفين

يجب التمييز بين حالتين مختلفتين:

الحالة I

نحسب

$$p = \frac{\text{عدد الوحدات في العيّنة من أي صف بمفرده}}{n} = \frac{a_1}{n}$$

أو

$$p = \frac{\text{العدد الكلي في العيّنة من مجموعة من الصفوف}}{n} = \frac{a_1+a_2+a_3}{n}$$

وعلى سبيل المثال، إذا صُنفت الأجوبة إلى «نعم»، «لا»، «لا أعرف»، و«لا جواب» فيمكن أخذ P مساوية لنسبة من أجابوا بنعم في العيّنة، أو بصورة أخرى نأخذ P مساوية لنسبة من أعطوا جوابًا محددًا (نعم أو لا) في العيّنة. وفي كلتا الحالتين، ومع أن التصنيف يحوي في الأصل أكثر من صفين، فإننا نحصل على قيمة P نفسها من خلال تقسيم الوحدات الـ n إلى صفين فقط. والنظرية التي قدمناها سابقًا تنطبق على هذه الحالة. ونحسب حدود الثقة وفق ما بيناه في الفقرة (٣-٦).

الحالة II

تُلغى أحيانًا صفوف معينة، وتُحسب p من تقسيم الصفوف الباقية إلى جزأين. وعلى سبيل المثال، يمكن إلغاء الأشخاص الذين لا يعرفون أو لم يقدموا جوابًا. ونأخذ بعين الاعتبار نسبة من قالوا «نعم» إلى مجموع من أجابوا بـ«نعم» أو «لا». والنسب التي تكون بُنيتها من هذا النوع هي في الغالب مثيرة للاهتمام في مسوح العيّنات. والمقام في نسبة كهذه ليس n، وإنما عدد ما n' أصغر من n.

ومع أن n' يتغير من عيّنة إلى أخرى فلا يزال من الممكن استخدام النتائج السابقة عن طريق أخذ التوزيع الشرطي لِـ p في عيّنات نثبّت فيها كلاًّ من n و n'. وقد لجأنا إلى هذه الحيلة في الفقرة (٢-١٢). لنفرض أن

$$p=\frac{a_1}{a_1+a_2}, \qquad n'=a_1+a_2, \qquad n=a_1+a_2+a_3$$

وبحيث يكون a_3 عدد الوحدات في العيّنة التي تنتمي إلى صفوف لا تحظى حاليًا باهتمامنا. فعندئذ، وكما هو مبين في الفقرة القادمة، يكون التوزيع الشرطي لِـ a_1 و a_2 هو التوزيع فوق الهندسي الذي نحصل عليه عندما يكون حجم العيّنة n' وحجم المجتمع $N'=A_1+A_2$ وبالتالي نجد من (3.19) أن التقريب الطبيعي لحدود الثقة الشرطية من أجل $P=A_1/(A_1+A_2)$ هو:

$$p\pm\left[t\sqrt{\left(1-\frac{n'}{N'}\right)\frac{pq}{(n'-1)}}+\frac{1}{2n'}\right] \qquad (3.22)$$

وإذا كانت قيمة N غير معروفة، فيمكن وضع n/N بدلاً من n'/N' في حد الـ ت م م في (3.22).

(٣-٩) التوزيع الشرطي لـ p

لإيجاد هذا التوزيع، نقصر انتباهنا على عيّنات حجمها n وفيها $n=a_1+a_2$ من الوحدات التي تنتمي إلى الصفين 1 و 2. وعدد العيّنات المتميزة من هذا النوع هو

$$\binom{N'}{n'}\binom{N-N'}{n-n'}=\binom{A_1+A_2}{a_1+a_2}\binom{A_3}{a_3} \quad (3.23)$$

ومن بين هذه العيّنات، نجد أن عدد تلك التي تتضمن a_1 من الصف 1 و a_2 من الصف 2 قد أعطي سابقًا في بسط العلاقة (3.21) ، فقرة (٣-٧). وبقسمة هذا البسط على (3.23) نجد

$$\Pr(a_1|A_1, A_2, n, n')=\binom{A_1}{a_1}\binom{A_2}{a_2}\Big/\binom{A_1+A_2}{a_1+a_2} \quad (3.24)$$

وهذا توزيع فوق هندسي عادي من أجل عيّنة حجمها n' من مجتمع حجمه $N'=A_1+A_2$.

مثال

ليكن المجتمع مؤلفًا من خمس وحدات b,c,d,e,f واقعة في ثلاثة صفوف

الصف	A_i	رموز الوحدات
1	1	b
2	2	c, d
3	2	e, f

وبأخذ عيّنات عشوائية حجمها 3 ، نرغب في تقدير $P=A_1/(A_1+A_2)$ أو، في هذه الحالة، 1/3 . وهكذا نجد $N=5$ و $N'=3$.

وتوجد 10 عيّنات ممكنة من الحجم 3 ، لكل منها الاحتمال نفسه، وسنصنّفها وفقًا لقيمة n' .

$n'=1$

العيّنة	a_1	a_2	p	الاحتمال الشرطي	$(p-P)$
bef	1	0	1	$\frac{1}{3}$	$\frac{2}{3}$
cef or *def*	0	1	0	$\frac{2}{3}$	$-\frac{1}{3}$

وإذا كانت العيّنات محددة بقيمتي a_1 و a_2 فيمكن الحصول على نوعين فقط: $a_1=1$ ، $a_2=0$ ؛ أو $a_1=0$ ، $a_2=1$ ، واحتمالاتها الشرطية 1/3 و 2/3 ، على الترتيب، تتفق مع العبارة العامة في (3.24) . وبالإضافة إلى ذلك

$$E(p)=\tfrac{1}{3}$$

$$\sigma_p^2=\left(\frac{1}{3}\right)\left(\frac{4}{9}\right)+\left(\frac{2}{3}\right)\left(\frac{1}{9}\right)=\frac{6}{27}=\frac{2}{9}$$

والتقدير p غير منحاز، ويتفق تباينه مع العلاقة العامة

$$\sigma_p^2=\left(\frac{N'-n'}{N'-1}\right)\frac{PQ}{n'}=\left(\frac{3-1}{3-1}\right)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)=\frac{2}{9}$$

وفي حالة $n'=2$ توجد ست عيّنات ممكنة، تعطي فقط مجموعتين من القيم لِـ a_1 ، a_2 .

$n'=2$

العيّنة	a_1	a_2	p	الاحتمال الشرطي	$(p-P)$
bce, bcf, bde, or *bdf*	1	1	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{6}$
cde or *cdf*	0	2	0	$\frac{1}{3}$	$-\frac{1}{3}$

والتقدير غير منحاز هنا أيضًا، وتباينه هو

$$\sigma_p^2=\left(\frac{2}{3}\right)\left(\frac{1}{36}\right)+\left(\frac{1}{3}\right)\left(\frac{1}{9}\right)=\frac{1}{18}$$

ويمكن التحقق من ذلك من العلاقة العامة. ونلاحظ أن التباين هو فقط 1/4 التباين الذي حصلنا عليه عندما كان $n'=1$. وفي معالجة شرطية يتغير التباين مع شكل العيّنة التي سحبناها.

وفي حالة $n'=3$ ، توجد فقط عيّنة واحدة ممكنة هي b,c,d . وهذا يعطي الكسر الصحيح للمجتمع 1/3 . والتباين الشرطي لِـ p هو صفر، كما تشير إلى ذلك العلاقة العامة التي تصبح صفرًا عندما $N'=n'$.

(٣-١٠) نسب ومجاميع فوق مجتمعات جزئية

إذا كنا نقوم بتقديرات منفصلة في كل من عدد من المجتمعات الجزئية أو ميادين الدراسة التي ستنتمي إليها وحدات العيّنة، فيمكن تطبيق نتائج الفقرتين (٣ - ٨) و(٣ - ٩). ويمكن تقديم بيان العيّنة كما يلي:

الصف	الميدان 1 C	C'	الميدان 2 C	C'	...	الميدان k C	C'	المجموع
عدد الوحدات	a_1	a_1'	a_2	a_2'	...	a_k	a_k'	n

ومن بين الوحدات الـ n ، وجدنا أن (a_1+a_1') تقع في الميدان 1 ومن بين هذه تقع a_1 في الصف C . ونقدر نسبة الوحدات الواقعة في الصف C من الميدان 1 بـ $p_1=a_1/(a_1+a_1')$: وقد نوقش التوزيع التكراري وحدود الثقة لـ p_1 في الحالة II من الفقرة (٣-٨)، وفي الفقرة (٣-٩). ولتقدير المجموع A_1 لوحدات الصف C في الميدان 1 ، هناك إمكانيتان، إذا كان N_1 ، العدد الكلي لوحدات الميدان 1 من المجتمع معروفًا، فيمكن استخدام التقدير الشرطي

$$\hat{A}_1 = N_1 p_1 = \frac{N_1 a_1}{a_1 + a_1'} \tag{3.25}$$

ونحسب خطأه المعياري على الشكل

$$s(\hat{A}_1) = N_1\sqrt{1-(n_1/N_1)}\sqrt{p_1 q_1/(n_1-1)} \tag{3.26}$$

حيث $n_1=a_1+a_1'$

وإذا كان N_1 غير معروف فالتقدير هو

$$\hat{A}_1' = \frac{N a_1}{n} \tag{3.27}$$

مع تقدير للخطأ المعياري هو

$$s(\hat{A}_1') = N\sqrt{1-(n/N)}\sqrt{pq/(n-1)} \tag{3.28}$$

حيث $p=a_1/n$.

(٣-١١) مقارنات بين ميادين مختلفة

بما أن تقديرات النسب تتم في الميادين المختلفة بصورة مستقلة، فيمكن القيام بمقارنات بين هذه النسب وفقًا للطرق الابتدائية المعتادة. وعى سبيل المثال، لاختيار ما إذا كانت النسبة $p_1=a_1/(a_1+a'_1)$ مختلفة بصورة مهمة عن النسبة $p_2=a_2/(a_2+a'_2)$ ، نشكل الجدول 2×2 المعتاد

	الميدان 1	2
C	a_1	a_2
C'	a_1'	a_2'
المجموع	n_1	n_1'

ويكون تطبيق اختبار χ^2 المعروف مناسباً [1958, Fisher] أو التقريب الطبيعي لتوزيع (p_1-p_2) . وبصورة مماثلة، نقوم بمقارنات بين النسب في أكثر من ميدانين بالطرق المعروفة في حالة جدول تصنيف $2\times k$.

ومن حين لآخر نرغب في اختيار ما إذا كان a_1 يختلف بصورة مهمة عن a_2 ، مثلاً، ما إذا كان عدد الجمهوريين الذين يفضلون مقترحاً ما أكبر من عدد الديمقراطيين الذين يفضلونه. وتحت الفرضية الابتدائية بأن هذين العددين متساويان في المجتمع، ينبغي أن ينقسم المجموع $n'=a_1+a_2$ ، في الميدانين المدروسين، بين الصفين باحتمالات متساوية. وبالتالي، يمكن اعتبار a_1 كعدد النجاحات في توزيع ثنائي يتضمن n' تكراراً، وباحتمال نجاح 1/2 وفقاً للفرضية الابتدائية. ويمكن التحقق من أن الانحراف الطبيعي (مصححاً من أجل الاستمرار) هو

$$\frac{2(|a_1-\frac{1}{2}n'|-\frac{1}{2})}{\sqrt{n'}}$$

(٣-١٢) تقدير النسب في المعاينة العنقودية

كما ذكرنا في الفقرة (٣-٢)، لا تصح الطرق السابقة إذا كانت كل وحدة عنقوداً من العناصر، بينما نقدِّر نحن نسبة العناصر الواقعة في صف C . وإذا كانت كل وحدة تتضمن العـدد نفسـه m من العناصر، وكانت $p_i = a_i / m$ نسبة العناصر في الوحدة i التي تنتمي إلى الصف C . فنسبة وحدات العيّنة الواقعة في C هي

$$p = \frac{\sum a_i}{nm} = \frac{1}{n}\sum^{n} p_i$$

أي أن التقدير p هو المتوسط غير المرجّح للكميات p_i . وبالتالي، إذا وضعنا p_i بدلاً من y_i ، أمكن تطبيق العلاقات في الفصل الثاني لتعطي التباين الحقيقي والتباين المقدَّر لِـ p .

$$V(p) = \frac{1-f}{n} \frac{\sum^{N} (p_i - P)^2}{N-1} \tag{3.29}$$

وتقدير العيّنة غير المنحاز لهذا التباين هو

$$v(p) = \frac{1-f}{n} \frac{\sum^{n} (p_i - p)^2}{n-1} \tag{3.30}$$

مثـال (١)

عولجت مجموعة من مرضى الجذام بدواء لمدة 48 أسبوعاً. ولقياس تأثير الدواء على عصيات الجـذام، اختبرنا جرثومياً وجود العصيات في ستة مواقع من جسم كل مريض. وبين الـ 366 موقعاً كان 153 ، أو 41.8% ، سلبياً. ما هو الخطأ المعياري لهذه النسبة؟ ويأتي هذا المثال من تجربة خاضعة لإرادة المجرب بدلاً من مسح إحصائي، إلا أنها توضح مدى الخطأ الذي نرتكبه باستخدام التوزيع الثنائي. من علاقة التوزيع الثنائي لدينا $n=366$ و

$$\text{s.e.}\,(p) = \sqrt{pq/(n-1)} = \sqrt{(41.8)(58.2)/365} = 2.58\%$$

وكل مريض هو عنقود من الوحدات فيه $m=6$ عناصر (مواقع) . ولإيجاد الخطأ المعياري باستخدام العلاقة الصحيحة، نحتاج إلى التوزيع التكراري للقيم الـ 61 لـ p_i . ومن الأسهل جدولة توزيع y_i عدد المواقع السلبية عند كل مريض. وبالتعبير عن p_i مئوياً يكون $p_i=100y_i/6$ ومن التوزيع في الجدول (٣-٤) نجد $\sum fy^2=669$ و

$$\text{s.e.}(\bar{y}) = \sqrt{\frac{\sum f_i(y_i-\bar{y})^2}{n(n-1)}} = \sqrt{\frac{669-[(153)^2/61]}{(61)(60)}} = 0.279$$

ومنه

$$\text{s.e.}(p) = \frac{100}{6}\text{s.e.}(\bar{y}) = 4.65\%$$

وهذا العدد هو حوالي 1.8 مرة القيمة التي يعطيها تطبيق العلاقة الثنائية. وتتطلب علاقة التوزيع الثنائي الفرض بأن النتائج من مواقع مختلفة من المريض نفسه مستقلة، مع أنه يوجد في الواقع ارتباط إيجابي قوي فيما بينها. ويبين السطر الأخير من الجدول (٣-٤) عدد المرضى المتوقع بـ ...,0,1,2 من المواقع السلبية، وهي محسوبة من التوزيع الثنائي $(0.58+0.42)^6$. لاحظ فرط التواترات الملحوظة f للمرضى الذين ليس لديهم مواقع سلبية أو لديهم خمسة أو ستة مواقع سلبية.

جدول (٣-٤) عدد المواقع السلبية للمريض الواحد

$y_i = 6p_i/100$	0	1	2	3	4	5	6	المجموع
f	17	11	4	4	7	14	4	61
fy_i	0	11	8	12	28	70	24	153
f_{exp}	2.3	10.1	18.3	17.6	9.6	2.8	0.3	61.0

وإذا كان حجم العنقود غير ثابت، وأخذنا m_i كعدد العناصر في الوحدة العنقودية i وكان $p_i=a_i/m_i$ فإن نسبة الوحدات في العيّنة الواقعة في الصف C هي

$$p = \frac{\sum^n a_i}{\sum^n m_i} \qquad (3.31)$$

ومن حيث تركيبها، نجد هنا نموذجاً للتقدير النسبة، وقد درسناه في الفقرة (٢-١١) وسندرسه في الفصل السادس وهو منحاز قليلاً، مع أنه من النادر أن يكون هذا الانحياز ذا أهمية عملية.

وإذا وضعنا a_i من أجل y_i و m_i من أجل x_i في (2.39)، نجد أن التباين التقريبي لـ p هو

$$V(p) \doteq \frac{1-f}{n\bar{M}^2} \frac{\sum (a_i - Pm_i)^2}{N-1} \tag{3.32}$$

حيث P هي نسبة عناصر الصف C في المجتمع و $\bar{M} = \sum^{N} m_i / N$ هو متوسط عدد العناصر في العنقود الواحد. والعبارة البديلة هي

$$V(p) \doteq \frac{1-f}{n} \sum^{N} \left(\frac{m_i}{\bar{M}}\right)^2 \frac{(p_i - P)^2}{N-1} \tag{3.33}$$

ويبين هذا الشكل أن التباين التقريبي يتضمن مجموع مربعات مرجَّحًا لانحرافات المقادير p_i عن القيمة p الخاصة بالمجتمع.

ومن أجل التباين المقدَّر نجد

$$v(p) = \frac{1-f}{n\bar{m}^2} \frac{\sum a_i^2 - 2p \sum a_i m_i + p^2 \sum m_i^2}{n-1} \tag{3.34}$$

حيث $\bar{m} = \sum m_i / n$ هو متوسط عدد العناصر للعنقود الواحد في العيّنة.

مثال (٢)

سُحبت عيّنة عشوائية بسيطة تتضمن 30 منزلاً من تعداد عام جرى سنة 1947 في الأحياء 6 و 7 من القطاع الصحي الشرقي من بَلتيمور. ويتضمن المجتمع حوالي 15000 منزل وفي الجدول (٣-٥) صنّفنا الأشخاص في كل منزل (ا) وفقاً لما إذا كانوا قد استشاروا طبيباً في الأشهر الاثني عشر الأخيرة، (ب) وفقاً للجنس.

وغايتنا هي مقارنة علاقة النسبة بعلاقة التوزيع الثنائي غير المناسبة. لنعتبر أولاً نسبة الأشخاص الذين استشاروا طبيباً. وسنأخذ في حالة التوزيع الثنائي

ومنه

$$n = 104, \qquad p = \frac{30}{104} = 0.2885$$

$$v_{bin}(p) = \frac{pq}{n} = \frac{(0.2885)(0.7115)}{104} = 0.00197$$

جدول (٣-٥) البيان الإحصائي لعيّنة عشوائية بسيطة من 30 منزلاً

رقم المنزل	عدد الأشخاص m_i	عدد: الذكور	عدد: الإناث a_i	مشاهدة الطبيب في العام الأخير: نعم a_i	مشاهدة الطبيب في العام الأخير: لا
1	5	1	4	5	0
2	6	3	3	0	6
3	3	1	2	2	1
4	3	1	2	3	0
5	2	1	1	0	2
6	3	1	2	0	3
7	3	1	2	0	3
8	3	1	2	0	3
9	4	2	2	0	4
10	4	3	1	0	4
11	3	2	1	0	3
12	2	1	1	0	2
13	7	3	4	0	7
14	4	3	1	4	0
15	3	2	1	1	2
16	5	3	2	2	3
17	4	3	1	0	4
18	4	3	1	0	4
19	3	2	1	1	2
20	3	1	2	3	0
21	4	1	3	2	2
22	3	2	1	0	3
23	3	2	1	0	3
24	1	0	1	0	1
25	2	1	1	2	0
26	4	3	1	2	2
27	3	1	2	0	3
28	4	2	2	2	2
29	2	1	1	0	2
30	4	2	2	1	3
المجموع	104	53	51	30	74

ومن أجل علاقة النسبة نلاحظ وجود 30 عنقوداً ونأخذ $n=30$

العدد الكلي في المنزل $m_i=i$

عدد الذين راجعوا الطبيب من المنزل $a_i=i$

كما سبق $p=0.2885$

$$\bar{m}=\frac{104}{30}=3.4667$$

$$\sum a_i^2=86; \sum m_i^2=404; \sum a_i m_i=113$$

ويمكن تجاهل الـ ت م م. وبالتالي نجد من (3.34)،

$$v(p)=\frac{(86)-2(0.2885)(113)+(0.2885)^2(404)}{(30)(29)(3.4667)^2}=0.00520$$

والتباين المعطى بطريقة النسبة، 0.00520، هو أكبر من ذلك المعطى بعلاقة التوزيع الثنائي، 0.00197. ولأسباب مختلفة، تختلف العائلات من حيث تواتر زيارة أفرادها للطبيب. وفي العيّنة ككل كانت نسبة الذين استشاروا طبيباً هي فقط أكثر من ربع بقليل، ولكن توجد بضع عائلات زار كل عضو فيها الطبيب. وقد نحصل على نتائج مشابهة لأية صفة يميل جميع الأفراد في الأسرة نفسها إلى أن يتصرفوا حيالها بالطريقة نفسها. وعند تقدير نسبة الذكور في المجتمع تكون النتائج مختلفة. وبحسابات من النوع نفسه نجد:

العلاقة الثنائية $v(p)=0.00240$

علاقة النسبة: $v(p)=0.00114$

وهنا تُغالي العلاقة الثنائية في تقدير التباين، والسبب يثير الاهتمام، فمعظم المنازل قد نشأت كنتيجة للزواج، وبالتالي فهي تتضمن ذكراً واحداً وأنثى واحدة، وبالتالي فإن نسبة الذكور للعائلة الواحدة تختلف بأقل من نصف ما هو متوقع من علاقة التوزيع الثنائي. ولا توجد أي من الأسر الثلاثين، باستثناء واحدة تتضمن فرداً واحداً، أي مؤلفة بكاملها من الذكور، أو بكاملها من الإناث. وإذا كان التوزيع الثنائي قابلاً للتطبيق بقيمة حقيقية لـ P تساوي النصف تقريباً، فستشكل المنازل التي يكون أفرادها من الجنس نفسه ربع المنازل التي حجمها 2، وثُمن المنازل التي

حجمها 4 . وقد ناقش Hansen و Hurwitz (1942) ، هذه الخاصة لنسبة الجنس . وأعطى Kish (1957) توضيحات أخرى للخطأ المرتكب نتيجة استخدام غير سليم لعلاقة التوزيع الثنائي في الأبحاث الاجتماعية .

تمارين

(٣-١) في مجتمع يتضمن $N=6$ ، $A=4$ ، $A'=2$ ، احسب قيمة a لكل العيّنات الممكنة التي حجمها 3 . تحقّق من النظريات المعطاة الخاصة بمتوسط وتباين $p=a/n$. وتحقّق أن

$$\frac{N-n}{(n-1)N}pq$$

هو تقدير غير منحاز لتباين p .

(٣-٢) في عيّنة عشوائية بسيطة حجمها 200 من مجتمع يحوي 2000 كلية ، كانت 120 كلية إلى جانب اقتراح معين، وكان 57 ضد الاقتراح، و 23 بدون رأي . قدّر 95% حدي ثقة لعدد الكليات في المجتمع التي توافق على الاقتراح .

(٣-٣) هل تعطي نتائج العيّنة السابقة دليلاً حاسماً على أن أغلبية الكليات في المجتمع توافق على الاقتراح؟

(٣-٤) يحتوي مجتمع فيه $N=7$ على العناصر B_1 ، C_1 ، C_2 ، C_3 ، D_1 ، D_2 و D_3 . وقد أُخذت عيّنة عشوائية بسيطة حجمها 4 لكي نقدر بواسطتها نسبة عدد العناصر C إلى عدد العناصر C و D . احسب التوزيع الشرطي لهذه النسبة p . وتحقق من صحة العلاقة الموافقة للتباين الشرطي .

(٣-٥) في المثال السابق، ما هو احتمال أن تحوي عيّنة حجمها 4 العنصر B_1 ؟ وبالتالي احسب التباين الوسطي لـ p فوق كل العيّنات العشوائية البسيطة التي حجمها 4 ، وتحقق من جوابك بوساطة العلاقة العامة .

290 أسرة من مدينة تحوي
(٣-٦) اختيرت عيّنة عشوائية بسيطة من
14828 أسرة. وقد سُئلت كل أسرة ما إذا كانت تملك أو تستأجر البيت، وما إذا كانت
أيضاً المستخدم الوحيد للمنافع الداخلية. وكانت النتائج كما يلي:

	المجموع	استئجاراً		ملكاً	
المستخدم الوحيد لدورة المياه		لا	نعم	لا	نعم
	290	34	109	6	141

(ا) قدّر من أجل الأسر المستأجرة نسبة من يستخدمون المنافع وحدهم واعط الخطأ المعياري لتقديرك.

(ب) قدّر العدد الكلي للعائلات المستأجرة في المدينة التي تشارك غيرها في المنافع واعط الخطأ المعياري لتقديرك.

(٣-٧) إذا كان العدد الكلي للعائلات المستأجرة في التمرين (٣-٦) في المدينة 7526 . قم بتقدير جديد لعدد المستأجرين الذين يشاركون غيرهم دورة المياه واعط الخطأ المعياري لهذا التقدير.

(٣ - ٨) لتقدير العدد الكلي للوحدات في الصف C في الميدان 1 (فقرة ٣-١٠)، يوصي بالتقدير $\hat{A}_1=N_1p_1$ إذا كان N_1 معروفاً وبـ $\hat{A}'_1=Na_1/n$ إذا كان N_1 مجهولاً. ومتجاهلاً الـ ت م م، بيّن أنه في العيّنات الكبيرة تكون نسبة تباين $\hat{A}_1$ إلى تباين $\hat{A}'_1$ مساوية $Q/(Q+p_1\pi)$ تقريباً، حيث النسبة من المجتمع غير الموجودة في الميدان 1 ، و P_1 كما في الفقرة (٣-١٠) هي نسبة الوحدات في الفترة 1 الواقعة في الصف C . أعرض الشروط التي تُنتج معها معرفة N_1 تخفيضاً كبيراً في التباين.

(٣-٩) في عيّنة عشوائية بسيطة حجمها 5 ، من مجتمع حجمه 30 ، لم توجد وحدات من الصف C في العيّنة. مستخدماً التوزيع فوق الهندسي، أوجد الحد الأعلى للعدد A من وحدات الصف C من المجتمع الموافق لـ 95% احتمال ثقة وحيدة الذيل.
أوجد أيضاً التقريب لـ A_U الذي نحصل عليه بحساب 95% حد ثقة أعلى P_U وتقصير

الفترة كما وصفناه في الفقرة (٣-٦). جرّب أيضاً الطريقة المذكورة على الصفحة ٨٦، مثال (٣).

(٣-١٠) لدى مصلحة الصحة الطلابية سجلّ بالعدد الإجمالي N للطلاب الذين يستحقون الخدمة الطبية وبالعدد الإجمالي Y للزيارات التي قام بها الطلاب خلال عام. ويعض الطلاب لم يقم بأية زيارة، وترغب المصلحة بتقدير العدد المتوسط للزيارات Y/N_1 من أجل الـN_1 طالباً الذين قاموا بزيارة واحدة على الأقل إلا أنها لا تعلم قيمة N_1. وقد أخذت عيّنة عشوائية بسيطة من n من الطلاب الذين يستحقون الخدمة الطبية. وتبيّن أن n_1 طالباً من الطلاب الـn في العيّنة قاموا بزيارة واحدة على الأقل وأن عدد الزيارات الكلي y، متجاهلين عامل الـ ت م م في هذا السؤال:

(ا) بيّن أن y/N_1 تقدير غير منحاز لـY/N_1 وأن تباينه الشرطي هو S^2/n_1. حيث S^2 تباين عدد الزيارات بين الطلاب الذين قاموا بزيارة واحدة على الأقل.

(ب) الطريقة الثانية لتقدير Y/N_1 هي أن تستخدم $\hat{N}_1=Nn_1/n$ كتقدير لـN_1 وبالتالي Yn/Nn_1 كتقدير لـY/N_1. بيّن أن هذا التقدير منحاز وأن نسبة الانحياز إلى القيمة الحقيقية Y/N_1 هي على وجه التقريب $(N-N_1)/nN_1$ أوجد عبارة تقريبية لتباين التقدير Y_n/Nn_1 وبيّن أن للتقدير في (ا) تبايناً أعلى إذا كان

$$S^2>\frac{(N-N_1)n_1}{N_1 n}\left(\frac{Y}{N_1}\right)^2$$

تلميح: إذا كان p تقديراً لـp وفقاً للتوزيع الثنائي، ومبنياً على n من التكرارات، فلدينا عندئذ وبصورة تقريبية

$$E\left(\frac{1}{p}\right)=\frac{1}{P}+\frac{Q}{nP^2},\qquad V\left(\frac{1}{p}\right)=\frac{Q}{nP^3}$$

(٣-١١) أي التقديرين السابقين يبدو أكثر دقة في الظروف الآتية؟ $N=2004$ ، $Y=3011$. بيّنت العيّنة، حيث $n=100$ ، أن 73 طالباً قاموا بزيارة واحدة على الأقل، وأن عدد الزيارات الكلي كان 152 وتقدير التباين s^2 هو 1.55 .

(٣-١٢) أخذنا عيّنة عشوائية بسيطة من n من الوحدات العنقودية، كل منها بـm من العناصر وذلك من مجتمع توجد فيه عناصر الصف C ، بنسبة p . عندما يكون الارتبـاط ضمن الـوحـدة العنقـودية متغـيراً فما هما أعـلى وأدنى قيمـة ممكنة للتباين الحقيقي لِـp (تقدير العيّنة لِـP) وكيف تكون بالمقارنة مع التباين محسوباً باستخدام التوزيع الثنائي؟ تجاهل عامل الـ ت م م.

(٣-١٣) في عينة الـ30 منزلاً في الجدول (٣-٥) يشير البيان الإحصائي المذكور أدناه إلى زيارات لطبيب الأسنان في العام الأخير. قدِّر تباين نسبة الأشخاص الذين رأوا طبيبًا، وقارِنه مع التقدير الثنائي للتباين.

زيارة طبيب الأسنان: لا	زيارة طبيب الأسنان: نعم	عدد الأشخاص	زيارة طبيب الأسنان: لا	زيارة طبيب الأسنان: نعم	عدد الأشخاص
4	1	5	4	1	5
0	4	4	6	0	6
3	1	4	2	1	3
2	1	3	1	2	3
3	0	3	2	0	2
3	1	4	3	0	3
3	0	3	2	1	3
2	1	3	2	1	3
1	0	1	3	1	4
2	0	2	4	0	4
4	0	4	2	1	3
2	1	3	2	0	2
3	1	4	5	2	7
2	0	2	3	1	4
4	0	4	3	0	3

(٣-١٤) إحـدى الـطرق لمعـاينة صفة نادرة، هي أن نستمر في سحب عيّنة عشـوائية بسيطة حتى نحصـل على m من الوحدات التي تمتلك هذه الصفة النادرة Haldane (1945) ، حيث يتقرر العدد m سلفاً. وإذا أمكن تجاهل عامل الـ ت م م، بيِّن أن احتمال أن يكون الحجم الإجمالي للعيّنة المطلوبة مساوياً n هو

$$\frac{(n-1)!}{(m-1)!(n-m)!}P^m Q^{n-m} \qquad (n \geq m)$$

حيث P هو تواتر الصفة النادرة. أوجد الحجم الوسطي لمحمل العيّنة وبيّن أنه إذا كان $m>1$ ، فإن $p=(m-1)/(n-1)$ هو تقدير غير منحاز لِـP. لمزيد من المناقشة انظر Finney (1949) و Sandelius (1951) ، الذي يتبنّى خطة تستمر فيها المعاينة حتى يتم العثور على m أو يبلغ الحجم الكلي للعينة الحد المقرر سلفاً n_0 . أيهما يحصل أولاً . انظر أيضاً الفقرة (٤-٥) .

تقدير حجم العيّنة

(٤-١) مثال افتراضي

عند تخطيط مسح إحصائي نصل دائماً إلى مرحلة يجب أن نتخذ فيها قراراً حول حجم العيّنة. وهذا القرار مهم. فعيّنة كبيرة جداً تتضمن هدراً للمصادر، وعيّنة صغيرة جداً تقللّ من فائدة النتائج. ولا يمكن اتخاذ القرار، دائماً، بصورة مُرضية، ذلك لأننا في الغالب لا نمتلك معلومات كافية تجعلنا تتأكد من أن اختيارنا لحجم العيّنة هو الاختيار الأفضل. وتقدم نظرية المعاينة إطاراً نفكر ضمنه تفكيراً منطقياً في المسألة المطروحة.

ويمكن أن يبينّ المثال الافتراضي الخطوات المطلوبة للوصول إلى حلّ. لنفرض أن مختصاً في الانثروبولوجيا يهيء لدراسة سكان جزيرة ما، ويرغب من بين أشياء أخرى، أن يقدِّر النسبة المئوية للسكان التي تمتلك الزمرة الدموية O . وقد تأمّن التعاون بحيث يصبح من الممكن أخذ عيّنة عشوائية بسيطة. فكم يجب أن يكون حجم العينة؟

لا يمكن مناقشة هذا السؤال قبل أن نتلقى أولاً جواباً عن سؤال آخر: ما هي الدقة التي يرغبها الاختصاصي في معرفته للنسبة المئوية للسكان الذين يمتلكون الزمرة الدموية؟ وفي الإجابة، يعرض الاختصاصي بأنه سيكون راضياً إذا كانت النسبة المئوية صحيحة في حدود 5± بالمائة، بمعنى أنه إذا كانت العيّنة تُظهر وجود 43 بالمائة من ذوي الزمرة الدموية O ، فإن النسبة المئوية في كامل الجزيرة تقع بالتأكيد بين 38 و 48 .

ولتجنب الالتباس، يكون من المستحسن أن نبيّن لذلك الاختصاصي أننا لا نتمكن من أن نضمن، بصورة مطلقة، دقة في حدود 5 بالمائة، إلا إذا قسنا الزمرة الدموية لكل فرد. إذ مهما كانت n كبيرة، فهناك فرصة للحصول على عيّنة غير محظوظة بالمرّة بحيث تقدم نتيجة خاطئة بأكثر من الدقة المرغوبة أي 5 بالمائة. ويجيب الاختصاصي ببرود أنه يعي هذا، وأنه يرحّب في أن يأخذ فرصة 1 من 20 في الحصول على عيّنة غير موفقة، وكل ما يطلبه هو قيمة n بدلاً من محاضرة في الإحصاء.

ونكون الآن في وضع نستطيع فيه القيام بتقدير تقريبي لِـ n. ولتبسيط الأمور، نتجاهل عامل الـ ت م م، ونفترض أن النسبة المئوية p في العيّنة تتوزع طبيعياً. ويمكن التحقق مما إذا كانت هذه الافتراضات منطقية عندما نعرف قيمة n بصورة مبدئية.

وبتعبير عملي، يجب أن تقع p ضمن المجال $(P\pm5)$، باستثناء فرصة واحدة من عشرين. وبما أننا نفترض أن p تتوزع طبيعياً حول P، فيجب أن تقع ضمن المجال $(P\pm2\sigma_p)$، باستثناء فرصة واحدة من عشرين. وبالإضافة إلى ذلك فإن:

$$\sigma_p \doteq \sqrt{PQ/n}$$

وهكذا يمكننا أن نكتب:

$$2\sqrt{PQ/n}=5 \qquad \text{أو} \qquad n=\frac{4PQ}{25}$$

وتبرز عند هذه النقطة صعوبة مشتركة بالنسبة لجميع المسائل المتعلقة بتقدير حجم عينة. فقد حصلنا على صيغة لِـ n ولكن n تعتمد على خاصيّة ما للمجتمع نقوم أصلاً بأخذ العيّنة لتقديرها. والخاصيّة، في هذا المثال، هي الكمية P التي نريد قياسها. ولذلك يجب أن نسأل الاختصاصي إذا كان قادراً على إعطائنا فكرة ما عن القيمة المحتملة لـ P. وبالاستناد إلى معلومات إحصائية سابقة حول جماعات عرقية أخرى، وإلى تخميناته المتعلقة بالتاريخ العرقي لهذه الجزيرة، يجيب بأنه سيكون مندهشاً إذا وقعت P خارج المجال 30 بالمائة إلى 60 بالمائة.

وهـذه المعلومـات كافية لتقـديم جواب مفيد. ذلـك لأنه من أجل أي قيمة لـ P بين 30 و 60 ، يقع الجداء PQ بين 2100 والنهاية العظمى لهذا الجداء 2500 التي تقع عند النقطة $P=50$. والقيمة n الموافقة تقع بين 336 و 400 . ولكي نكون على الجانب الآمن، نأخذ القيمة 400 قيمة مبدئية لحجم العينة n .

ويمكن الآن إعـادة النـظر في الفـرضيات التي اعتمـدناها في هذا التحليل. ومع $n=400$ و P بين 60 و 30 ، ينبغي أن يكون توزيع p قريباً من التوزيع الطبيعي. وما إذا كان عامل الـ ت م م مطلوباً أم لا يعتمد على عدد سكان الجزيرة. وإذا تجاوز المجتمـع 8000 ، فإن كسر المعاينة أقل من 5 بالمائة وليس من الضروري القيام بأي تعديل من أجل عامل الـ ت م م. ونناقش في الفقرة (٤-٤) طريقة تطبيق التعديل المطلوب من جديد، في حال الحاجة إليه.

(٤ ـ ٢) تحليل المسألة

الخطوات الرئيسة التي يتضمنها اختيار العيّنة هي كما يلي:

١ ـ يجب أن يكون هناك تصور ما حول ما نتوقعه من العيّنة. ويمكن أن يكون هذا التصـور بدلالـة حدود الخطأ المرغوبة، كما في المثال السابق، أو بدلالة قرار ما سنتخذه أو عمل سنقوم به عندما تصبح نتائج العينة معروفة. وتبقى مسؤولية تأطير التصور، بصورة رئيسة، على عاتق الأشخاص الذين يرغبون في استخدام نتائج المسح الإحصائي، علماً بأنهم يحتاجون، في الغالب، للإرشاد كي توضع رغباتهم في شكل عددي.

٢ ـ يجب إيجاد معادلة ما تربط n بالدقة المرغوبة للعيّنة. وستتغير المعادلة وفقاً لمحتوى الدقة المرغوبة، ووفقاً لنوع المعاينة الذي سيجري تطبيقه. وإحدى فوائد المعاينة الاحتمالية هي أنها تمكّننا من وضع مثل هذه المعادلة.

٣ ـ ستحوي هذه المعادلة بعض الخواص المجهولة للمجتمع على شكل معالم.

٤ ـ ويحدث غالباً أن تُنشر معلومات إحصائية تتعلق بأجزاء رئيسة معيّنة من المجتمع، وتوضع حدود الخطأ المرغوبة لكل جزء ونقوم بحسابات منفصلة لقيمة n في كل جزء، ثم نحسب الحجم الكلي n بالجمع.

٥ ـ تقاس عادة أكثر من مفردة أو خاصة في مسح معاينة، ويكون عدد المفردات أحياناً كبيراً. وإذا وضعنا الدرجة المرغوبة من الدقة لكل مفردة، فقد تقود الحسابات إلى سلسلة من القيم المتعارضة لِـ n، واحدة لكل مفردة، ويجب إيجاد طريقة ما للتوفيق بين هذه القيم.

٦ ـ وأخيراً يجب تثمين القيمة المختارة لِـ n لرؤية ما إذا كانت تتلاءم مع المصادر المتوافرة لأخذ العيّنة. وهذا يتطلب تقديراً للتكلفة والعمل والوقت، والمواد المطلوبة للحصول على عيّنة الحجم المقترح. ويصبح بادياً للعيان أحياناً أنه لابدّ من تخفيض كبير في قيمة n. ولابد عندئذ من مواجهة قرار صعب ـ فإما أن نمضي بعيّنه ذات حجم أصغر بكثير، وبالتالي نخفض الدقة، أو أن نهجر المشروع حتى تتوافر لنا موارد أكثر.

وسنناقش، في الفقرات القادمة، بعض هذه المسائل بتفصيل أكبر.

(٤ ـ ٣) تحديد الدقة

يمكن التعبير عن الدقة المرغوبة بعرض كمية الخطأ في تقديرات العيّنة التي نقبل التسامح بها. ونُحدّد هذه الكمية، بأفضل شكل نستطيعه، في ضوء الاستخدامات المنتظرة لنتائج العيّنة. ويكون من الصعب أحياناً أن نقرر مدى الخطأ الذي ينبغي التسامح به، خاصة عندما يكون للنتائج استخدامات عديدة مختلفة. لنفرض أننا سألنا الاختصاصي في علم الأنثروبولوجيا عن السبب الذي جعله يرغب في أن تكون النسبة المئوية لذوي الزمرة الدموية O صحيحة في حدود 5 بالمائة، وليس مثلاً، 4 أو 6 بالمائة. فقد يجيب بأن الاستخدام الرئيس للبيان الإحصائي للزمرة الدموية هو في مجال التصنيف العنصري. ويظن بقوة أن انتماء سكان الجزيرة هو إما إلى نوع عنصري تكون النسبة المئوية فيه حوالي 53 بالمائة، أو إلى نوع تكون النسبة المئوية فيه حوالي 50 بالمائة. ويبدو له أن حداً للخطأ في التقدير مقداره خمسة في المائة هو صغير بكفاية كي يسمح له بالتصنيف إلى أحد هذين النوعين. وعلى أي حال فليس لديه اعتراض شديد على أن يكون حد الخطأ 4 أو 6 بالمائة.

وهكذا فإن اختيار الـ 5 بالمائة من قبل الاختصاصي كان إلى حد ما، اختيارًا كيفياً. وفي هذا المجال يقدم المثال السابق نموذجاً عن الطريقة التي نقرر فيها غالباً حول حد للخطأ. وفي الحقيقة كان ذلك الاختصاصي أكثر إحاطة بما يريد مما سيوجد عند الكثير من العلماء الآخرين أو الكثير من الإداريين. وعند إثارة مسألة الدرجة المرغوبة من الدقة للمرة الأولى، فقد يعترف مثل هؤلاء الأشخاص بأنهم لم يفكروا قطّ بالمسألة، وليس لديهم أية أفكار حول الجواب. وعلى أي حال، فقد كانت خبرتي أنه بعد نقاش بسيط، سيتمكنون في الغالب من الإشارة، ولو بصورة تقريبية على الأقل، لحجم حد الخطأ الذي يبدو معقولاً بالنسبة لهم.

وقد لا نستطيع في العديد من الحالات التطبيقية المضي لأبعد من ذلك. ويقع جزء من الصعوبة في أننا لا نعرف ما يكفي عن تبِعات حجوم مختلفة للخطأ من حيث تأثيرها على حكمة وصحة القرارات التطبيقية التي ستُتخذ بوحي من نتائج المسح الإحصائي. ويستحق هذا الموضوع من الدراسة أكثر مما يتلقاه حاليًا. ومع تجمّع المعرفة وتراكمها سيصبح اختيار الدرجة المرغوبة في الدقة أسهل. وحتى عندما تكون تَبعات الأخطاء معروفة فهناك، على أي حال، العديد من المسوح الإحصائية المهمة التي ستُستخدَم نتائجها من قِبل أناس مختلفين ولغايات مختلفة، وبعض هذه الغايات لا تكون مرئية عند تخطيط المسح الإحصائي. وبالتالي فإن عنصراً من العمل التخميني قد يفرض نفسه عند تحديد الدقة وذلك لفترة لاحقة من الزمن.

وإذا أُخذت العينة لهدف هو في غاية التحديد، وليكن مثلاً اتخاذ قرار نعم أو لا، أو لتقرير مبلغ المال الذي سينفق على مغامرة معيّنة، فيمكننا، عادة، عرض الدقة التي نحتاجها بطريقة أكثر حسماً، وذلك بدلالة تبِعات الخطأ في القرار. ونعطي معالجة عامة لمسائل من هذا النوع في الفقرة (٤-٨)، وهي تقدم، بالرغم من الحاجة إلى التوسع فيها، بداية منطقية لحلّ.

(٤-٤) قانون يتعلق بـ n عند معاينة النسب

تُصنّف الوحدات إلى صفين C و C'، ونتم الموافقة على هامش ما d للخطأ في النسبة المقدّرة لوحدات الصف C وهي P، وأننا نقبل التعرض لمخاطرة تجاوز الخطأ الفعلي للمقدار d، ولكن باحتمال صغير α، أي أننا نريد

$$Pr(|p-P| \geq d)=\alpha$$

ونفترض معاينة عشوائية بسيطة، كما نفترض أن p تتوزع طبيعيًّا. ومن النظرية (٣-٢) في الفقرة (٣-٢) نجد

$$\sigma_p=\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{PQ}{n}}$$

ومن ثمّ فإن العلاقة التي تربط n بالدرجة المرغوبة للدقة هي

$$d=t\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{PQ}{n}}$$

حيث t هي قيمـة المتغير الطبيعي المعياري التي تقطع مساحة قدرها $\frac{\alpha}{2}$ من كل من الذيلين. وبحلها من أجل n، نجد:

$$n=\frac{\dfrac{t^2PQ}{d^2}}{1+\dfrac{1}{N}\left(\dfrac{t^2PQ}{d^2}-1\right)} \qquad (4.1)$$

وفي التطبيق العملي، نبدل P في هذا القانون بتقدير سلفي p. وإذا كان N كبيراً فإن التقريب الأول يعطي

$$n_0=\frac{t^2pq}{d^2}=\frac{pq}{V} \qquad (4.2)$$

حيث:

$$V=\frac{pq}{n_0}= \text{التباين المرغوب لنسبة العيّنة}$$

وعند التطبيق نحسب أولاً n_0 . فإذا كان n_0/N مهملاً، نعتبر n_0 تقريباً مرضياً لـ n في المعادلة (4.1) . وإذا لم يتوافر ذلك، فمن الظاهر من مقارنة (4.1) و (4.2) أننا نحصل على n من العلاقة :

$$n = \frac{n_0}{1+(n_0-1)/N} \doteq \frac{n_0}{1+(n_0/N)} \qquad (4.3)$$

مثال :

في المثال الافتراضي لزمر الدم، لدينا :

$$d = 0.05, p = 0.5, \alpha = 0.05, t = 2$$

ومنه :

$$n_0 = \frac{(4)(0.5)(0.5)}{(0.0025)} = 400$$

لنفرض أنه يوجد 3200 شخص فقط على الجزيرة. فنحتاج لعامل الـ ت م م، ونجد عندئذ أن

$$n = \frac{n_0}{1+(n_0-1)/N} = \frac{400}{1+\frac{399}{3200}} = 356$$

وإذا عُبّر عن كل من d و p و q كنسب مئوية بدلاً من نسب في صيغة n_0 ، فإن الصيغة تبقى صحيحة. وبما أن الجداء pq يتزايد عندما تتحرك في اتجاه الـ $\frac{1}{2}$ أو 50 بالمائة، فإننا نحصل على تقدير محافظ لِـ n باختيارنا قيمة لِـ p هي، ضمن المدى الذي نعتقد أنه من المرجح أن يحوي p ، أقرب ما يمكن إلى النصف. وإذا بدا أنه من المرجح أن تقع p ، مثلاً، بين 5 بالمائة و9 بالمائة، فنفترض لتقدير n أن $p=0.09$.

وأحياناً، وبصورة خاصة عند تقدير العدد الكلي NP لوحدات الصف C ، فإننا نرغب في التحكم بالخطأ النسبي r بدلاً من الخطأ المطلق في NP ، وعلى سبيل المثال، قد نرغب في تقدير NP بخطأ لا يتجاوز 10% أي أننا نريد

$$\Pr\left(\frac{|Np-NP|}{NP} \geq r\right) = \Pr(|p-P| \geq rP) = \alpha$$

ومن أجل هذا التحديد، نعوض rP أو rp بدلاً من d في العلاقتين (4.1) و (4.2)
ومن (4.2) نجد :

$$n_0 = \frac{t^2 pq}{r^2 p^2} = \frac{t^2}{r^2} \frac{q}{p} \qquad (4.2)'$$

ولا تتغير العلاقة (4.3) .

(٤ ـ ٥) المفردات النادرة ـ المعاينة العكسية

عند تقدير n من العلاقات (4.1) ، (4.2) ، يعوض المعاين أفضل تقدير مسبق يعتقده لنسبة المجتمع P . وإذا كان معروفاً أن P تقع بين 30% و 70% ، كما في المثال في الفقرة (٤-١)، فلا يشكل التقدير الدقيق لِـ P أمراً حاسماً. ولكن في مفردة نادرة (مثلاً $p \leqslant 10\%$)، يكون الحجم n الضروري من أجل خطأ نسبي محدد r أكبر بإحدى عشرة مرة عند $P=1\%$ مما هو عند $P=10\%$. وفي هذه الحالة (P صغيرة إلا أنها ليست معروفة جيداً بصورة مُسبقة)، يكون لطريقة Haldane (1945) ، حيث نستمر في المعاينة حتى نعثر على m من المفردات النادرة، ميزة مهمة. وتُدعى هذه الطريقة عادة المعاينة العكسية.

وإذا كان n حجم العيّنة الذي يظهر عنده المفرد النادر الـ m ، $(m>1)$ ، فإن تقديراً غير منحاز لِـ P هو $p=(m-1)/(n-1)$. ومن أجل N كبير جدًا و P صغيرة و $m \geqslant 10$ ، يمكن البرهان أن، $mP^2Q/(m-1)^2$ هو تقدير جيد لِـ $V(p)$. وبالتالي،

$cv(p)=(mQ)^{1/2}/(m-1)<\sqrt{m}/(m-)$ ، وسيكون هذا حداً أعلى قريباً إذا كانت P صغيرة. وهكذا، ومن خلال تثبيت m سلفاً، يمكن التحكم بقيمة $cv(p)$ دون معرفة مسبقة بِـ P . والقيمة $m=27$ تعطي $cv(p)<20\%$ ، إلا أننا نحتاج إلى $m=102$ من أجل $cv(p)<10\%$. وقيمة n في هذه الطريقة هي متغير عشوائي ، وستكون قيمته كبيرة إذا كانت P صغيرة.

(٤ - ٦) العلاقة الخاصة بـ n في حالة بيان إحصائي من طبيعة مستمرة

في معظم الأحيان، نرغب في التحكم بالخطأ النسبي r في تقديرات مجموع أو متوسط مجتمع. ومع عيّنة عشوائية بسيطة متوسطها $\bar{y}$ ، نريد:

$$\Pr\left(\left|\frac{\bar{y}-\bar{Y}}{\bar{Y}}\right|\geq r\right)=\Pr\left(\left|\frac{N\bar{y}-N\bar{Y}}{N\bar{Y}}\right|\geq r\right)=\Pr\left(|\bar{y}-\bar{Y}|\geq r\bar{Y}\right)=\alpha$$

حيث α احتمال صغير. ونفترض أن $\bar{y}$ يتوزع بصورة طبيعية، ومن النظرية (٢-٢)، نتيجة (١) يكون خطؤه المعياري

$$\sigma_{\bar{y}}=\sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}}$$

وبالتالي

$$r\bar{Y}=t\sigma_{\bar{y}}=t\sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}} \qquad (4.4)$$

وبحلها من أجل n نجد

$$n=\left(\frac{tS}{r\bar{Y}}\right)^2\Big/\left[1+\frac{1}{N}\left(\frac{tS}{r\bar{Y}}\right)^2\right]$$

ونلاحظ أن خاصة المجتمع التي يعتمد عليها n هي معامل اختلافه $S/\bar{Y}$. وغالباً ما يكون أكثر استقراراً من S وتخمينه سلفاً أسهل من تخمين S نفسها.

وكتقريب أول نأخذ

$$n_0=\left(\frac{tS}{r\bar{Y}}\right)^2=\frac{1}{C}\left(\frac{S}{\bar{Y}}\right)^2 \qquad (4.5)$$

وبتعويض تقدير مسبق لِـ $(S/\bar{y})$ يكون المقدار C هو مربع معامل الاختلاف المرغوب لتقدير العيّنة. وإذا كانت n_0/N كبيرة نحسب n كما في العلاقة (4.3) وهي

$$n=\frac{n_0}{1+(n_0/N)}$$

وإذا أردنا التحكم بـ d ، الخطأ المطلق في $\bar{y}$ ، بدلاً من الخطأ النسبي r نأخذ $n_0 = t^2S^2/d^2 = S^2/V$ حيث V التباين المرغوب لـ $\bar{Y}$.

مثال

في المشاتل التي تُنتج أغراساً جديدة للبيع ، يُستحسن ، في أواخر الشتاء أو أوائل الربيع ، تقدير عدد الأغراس السليمة التي نتوقع أن تكون تحت تصرفنا . لأن هذا يحدد سياستنا تجاه قبول الطلبات التجارية وتجاه الالتماسات الخاصة . وقد قام Johnson (1943) بدراسة لطرق المعاينة بغية تقدير العدد الكلي للأغراس . وقد حصل على المعلومات الإحصائية التالية من مسكبة لأغراس القيقب (silver maple) ، عرضها قدم واحد ، وطولها 430 قدماً . وكانت وحدة المعاينة قدماً في اتجاه طول المسكبة ، أي أن $N=430$. وبالتعداد الكامل للمسكبة ، وُجد أن $\bar{Y}=19$ ، $S^2=85.6$ وهي القيم الصحيحة للمجتمع .

وفي حالة معاينة عشوائية بسيطة ، كم يجب أن نأخذ من وحدات المعاينة لتقدير $\bar{Y}$ في حدود 10 بالمائة ، وذلك باستثناء فرصة من عشرين؟ ومن المعادلة (4.5) نحصل على

$$n_0 = \frac{t^2S^2}{r^2\bar{Y}^2} = \frac{(4)(85.6)}{(1.9)^2} = 95$$

وبما أن n_0/N ليس مهملاً ، نأخذ

$$n = \frac{95}{1 + \frac{95}{430}} = 78$$

أي أننا نضطر لتعداد 20 بالمائة تقريباً من كامل المسكبة حتى نبلغ الدقة المطلوبة .

والعلاقات المعطاة هنا من أجل n تنطبق فقط على معاينة عشوائية بسيطة نستخدم فيها متوسط العيّنة تقديرًا لـ $\bar{Y}$. والعلاقات المناسبة لطرق أخرى في المعاينة والتقدير سنقدمها عند مناقشة تلك الطرق .

(٤-٧) تقديرات مسبقة لتباين مجتمع

لا يشكل مثال المشاتل مثالاً نموذجياً من حيث إن تباين المجتمع S^2 كان معروفاً. وعمليًا هناك أربع طرق تقدير للوصول إلى تقدير لتباين المجتمع نستخدمه في تحديد حجم العيّنة: (١) بأخذ العيّنة على مرحلتين، في المرحلة الأولى نأخذ عينة عشوائية بسيطة حجمها n_1 ونحصل منها على S_1^2 تقديرًا لـ S^2 أو p_1 تقديرًا لـ P كما نحصل على الحجم المطلوب n، (٢) باستخدام نتائج مسح إحصائي استكشافي، (٣) بوساطة معاينة سابقة من المجتمع نفسه أو من مجتمع مشابه، و(٤) بعملية تخمين حول بنية المجتمع مع الاستعانة ببعض النتائج الرياضية.

وتعطي الطريقة (١) التقديرات الأكثر جدارة بالثقة لـ S^2 أو P، إلا أنها لا تُستخدم كثيراً باعتبارها تبطىء عملية إتمام المسح الإحصائي. وعندما تكون الطريقة ملائمة، يبيّن Cox (1952) متبعاً عملاً قام به Stein (1945)، كيفية حساب n من A_1^2 أو p_1 بحيث يكون للتقدير النهائي $\bar{y}$ أو p تباين محدد سلفاً V، أو حدٌّ للخطأ d محدد سلفاً، أو معامل اختلاف محدد سلفاً. ونفترض أن العيّنة الأولى كبيرة بدرجة يمكننا معها إهمال حدود من مرتبة $\frac{1}{n_1^2}$. ونقتبس هنا قليلاً من النتائج.

وتفترض النتائج المعطاة هنا أن $n_1 \leqslant n$ حيث n حجم العيّنة النهائي. وعندما لا يكون الأمر كذلك يمكن العودة إلى Cox (1952).

تقدير $\bar{Y}$ علماً أن $cv=\sqrt{C}$

تفترض النتائج أن y_i يتوزع وفق التوزيع الطبيعي. وإذا كان S_1^2 التباين المقدّر من العيّنة الأولى، نأخذ وحدات إضافية لجعل الحجم النهائي للعيّنة

$$n=\frac{s_1^2}{C\bar{y}_1^2}\left(1+8C+\frac{s_1^2}{n_1\bar{y}_1^2}+\frac{2}{n_1}\right) \tag{4.6}$$

ويكون المتوسط $\bar{y}$ للعيّنة النهائية منحازاً قليلاً. خذ $\hat{\bar{Y}}=\bar{y}(1-2C)$.

تقدير $\bar{Y}$ بتباين يساوي V

خذ وحدات إضافية لجعل حجم العيّنة النهائي

$$n=\frac{s_1^2}{V}\left(1+\frac{2}{n_1}\right)$$

وإذا كان S معروفاً بالضبط فسيكون حجم العيّنة المطلوب S^2/V . وتأثير عدم معرفة S هو زيادة متوسط الحجم بنسبة $\left(1+\frac{2}{n_1}\right)$.

تقدير P بتباين يساوي V

ليكن p_1 تقدير P من العيّنة الأولى. فالحجم المركب للعيّنتين الأولى والثانية ينبغي أن يكون

$$n=\frac{p_1q_1}{V}+\frac{3-8p_1q_1}{p_1q_1}+\frac{1-3p_1q_1}{Vn_1} \tag{4.8}$$

والحد الأول من الطرف الأيمن هو الحجم المطلوب إذا كنا نعلم أن $P=p_1$. وبهذه الطريقة، يكون التقدير الثنائي العادي p الناتج عن العيّنة ذات الحجم n بكاملها منحازاً قليلاً. ولتصحيح هذا الانحياز، نأخذ

$$\hat{P}=p+\frac{V(1-2p)}{pq}$$

تقدير P حيث $cv=\sqrt{C}$ معطى

خذ

$$n=\frac{q_1}{Cp_1}+\frac{3}{p_1q_1}+\frac{1}{Cp_1n_1} \tag{4.9}$$

والتقدير هو $\hat{P}=p=Cp/q$. وفي جميع النتائج المذكورة أعلاه تجاهلنا عامل الـ ت م م.

مثال

يرغب معاين في تقدير P بمعامل اختلاف 0.1 (10%) . ويُظن أن P ستقع في مكان ما بين 5 و 20% . وهذا المدى عريض إلى درجة لا تسمح بإعطاء تقدير أولي جيد للحجم n المطلوب، وبما أن cv للنسبة P هي $\sqrt{Q/nP}$. فمن السهل التحقق

من أن $n=400$ مناسب لِـ $P=20\%$. إلا أننا سنحتاج إلى $n=2900$ إذا كان P مساوياً 5% فقط.

ولـذلـك، يأخـذ المعـاين عيّنـة ابتـدائية فيهـا $n_1=396$ ويجـد $p_1=0.101$. وبما أن $\sqrt{C}=0.1$ ، $C=0.01$ فالمعادلة (4.9) تعطي

$$n=\frac{(0.899)}{(0.01)(0.101)}+\frac{3}{(0.0908)}+\frac{1}{(0.01)(40)}=926$$

والعيّنة المركبة تعطي $np=88$ ، $p=88/926=0.0950$. ويبلغ تصحيح الانحياز Cp/q ، 0.0011 مما يؤدي إلى تقدير نهائي 0.094 أو 9.4% .

والطريقة الثانية هي القيام بمسح استكشافي صغير يخدم أهدافاً عدة، وبصورة خاصة إذا كانت إمكانية القيام بالمسح الرئيس موضع شكّ. وإذا كان المسح الاستكشافي نفسه عيّنة عشوائية بسيطة فيمكن تطبيق الطرق السابقة. ولكن غالباً ما يكون العمل الاستكشافي مقتصراً على جزء من المجتمع يسهُل علينا تناوله أو أنه يهدف للكشف عن حجم بعض المشاكل. ولابد أن نضع في حسابنا الطبيعة الاختيارية للاستكشاف عند استخدام نتائجه لتقدير S^2 أو P . وعلى سبيل المثال، من المتعارف عليه عملياً أن يقتصر العمل الاستكشافي على عناقيد قليلة من الوحدات. وهكذا يقيس s^2 الذي نحسبه أكثر ما يقيس التغير ضمن عنقود وقد يكون تقديراً بالنقصان لِـ S^2 . وسنناقش في الفصل التاسع العلاقة بين تغيّر ما ضمن العنقود والتغيّر ما بين العناقيد. وتبرز المسألة نفسها في المعاينة العنقودية الخاصة بالنسب، وفيها يمكن أن تكون الصيغة pq/n تقديراً بالنقصان لتأثير التغير بين العناقيد. ويعطي Cornfield (1957) توضيحاً جيداً لتقدير حجم عيّنة في معاينة عنقودية تتعلق بالنسب.

الطريقة الثالثة: استخدام نتائج مسوح إحصائية سابقة - وتشير إلى فائدة نشر أيّة معلومات حول الانحرافات المعيارية، تمّ الحصول عليها من مسوح سابقة، أو جعلها، على الأقل، في متناول اليد لمن يحتاجها. ومن سوء الحظ فإن تكلفة حساب

في مسوح معقدة هي تكلفة مرتفعة حتى في حالة استخدام الانحرافات المعيارية في مسوح معقدة هي تكلفة مرتفعة حتى في حالة استخدام الحاسب الآلي، وفي معظم الحالات نحسب ونسجّل فقط التقديرات الرئيسة، وإذا عثرنا على معلومات مناسبة من الماضي فقد تتطلب قيمة S^2 تعديلاً يأخذ في الاعتبار تغيرات الزمن. وفي حالة بيانات غير متناظرة وحيث يتغير $\bar{Y}$ فيها مع الزمن، نجد في الغالب أن S^2 يتغير وفق نسبة تقع في مكان ما بين $k\bar{Y}, k\bar{Y}^2$، حيث k عدد ثابت. وهكذا إذا بدا لنا أن $\bar{Y}$ قد ازداد بنسبة 10% خلال الفترة الزمنية التي انقضت منذ قيام المسح السابق، فيمكن زيادة تقديرنا الابتدائي لِـ S^2 بِـ 10 إلى 20%.

وأخيراً، يمكن أحياناً الوصول إلى تقدير مفيد لِـ S^2 من معلومات قليلة نسبياً حول طبيعة المجتمع. وفي دراسات قديمة حول عدد الدود الشريطي في التربة، استُخدمت أداة تأخذ عيّنة صغيرة من التربة السطحية (5×9×5 بوصة) ولتقدير حجم العيّنة n يحتاج المعاين لمعرفة الانحراف المعياري لعدد الدود الشريطي الذي يمكن أن يحويه تجويف في أداة المعاينة. وإذا كان توزيع الدود الشريطي عشوائياً فوق التربة السطحية فقد يتبع العدد الموجود في حجم صغير من التربة توزيع بواسون الذي يتميز بأن تباينه S^2 يساوي متوسطه $\bar{Y}$. وبما أن الدود الشريطي ينزع الى التكوّم، فقد تقرر افتراض $S^2=1.2\bar{Y}$ والعامل 1.2 هو عامل أمان كيفي. ومع أن $\bar{Y}$ نفسه، لم يكن معروفاً فقد أمكن وضع مخطط لقيم $\bar{Y}$ ذات الأهمية الاقتصادية وذلك من حيث صلتها بتلف المحصول الزراعي. وهاتان القطعتان من المعلومات سمحتا بتحديد حجوم للعيّنة ثبتت جودتها.

وبيّن Deming (1960) كيف يمكن استخدام بعض التوزيعات الرياضية البسيطة لتقدير S^2 وذلك من معرفة مدى التوزيع وفكرة عامة عن شكله. وإذا كان التوزيع كالتوزيع الثنائي، بنسبة p من الملاحظات في إحدى نهايتي المدى ونسبة q في النهاية الأخرى، فعندئذ $S^2=pqh^2$ حيث h هو المدى. وكعلاقات أخرى مفيدة نذكر $S^2=0.083h^2$ في حالة توزيع مستطيل، $S^2=0.056h^2$ في حالة توزيع له شكل المثلث القائم، و $S^2=0.0422h^2$ في حالة مثلث متساوي الساقين.

ولا تقدم هذه العلاقات الكثير من العون إذا كان h كبيراً أو أن معرفتنا به هي معرفة تنقصها الدقة. وعلى أي حال، إذا كان h كبيراً، فمن الجيّد، كأسلوب معاينة، أن نلجأ إلى تقسيم المجتمع إلى طبقات (فصل ٥) بحيث ينخفض المدى ضمن كل طبقة. وعادة يصبح الشكل أبسط أيضاً (أقرب إلى شكل التوزيع المستطيل) ضمن طبقة واحدة. وهكذا تكون هذه العلاقات فعالة فيما يتعلق بالتنبؤ بـ S^2 وبالتالي التنبؤ بـ n ضمن كل طبقة بمفردها.

(٤ ـ ٨) حجم العيّنة في حالة أكثر من مفردة واحدة

في معظم المسوح الإحصائية، نجمع معلومات حول أكثر من مفردة واحدة. وإحدى الطرق لتحديد حجم عيّنة هو أن نحدد هوامش الخطأ لتلك المفردات التي نعتبرها أكثر أهمية في المسح الإحصائي. ونقوم أولاً بتحديد حجم العيّنة الذي نحتاجه لكل من المفردات المهمة بمفردها.

وعندما ننتهي من تقدير n لمفردة واحدة، يحين الوقت لكي نقوّم الحالة. وقد يحدث أن تكون الحجوم n المطلوبة كلها قريبة من بعضها بصورة معقولة. وإذا وقع أكبر المقادير n ضمن حدود الميزانية، نختار هذه القيمة لـ n. وبصورة عامة، يوجد تغير كافٍ بين المقادير n المطلوبة بحيث نتردد في اختيار الأكبر، إما لاعتبارات تتعلق بالميزانية أو بسبب أن مثل هذا الاختيار سيعطي مقياس دقة إجماليا أعلى بكثير مما فكّرنا فيه أصلاً. وربما أمكن في مثل هذه الحالة التساهل في معيار الدقة المرغوب لبعض المفردات، مما يسمح لنا باستخدام قيمة أصغر لـ n.

وفي بعض الحالات تكون قيم n المرغوبة لمفردات مختلفة، من التعارض بحيث يتوجب حذف بعض المفردات من حدود المصادر المتوافرة. وقد لاتكون الصعوبة بسبب حجم العيّنة فقط، إذ قد تستدعي بعض المفردات نوعاً من المعاينة يختلف عما تستدعيه مفردات أخرى. ومن المفيد، في مجتمعات أخذنا منها عيّنات بصورة متكررة، أن نجمّع معلومات حول تلك المفردات التي يمكن، من وجهة النظر الاقتصادية، ضمّها في

مسح إحصائي عام، وتلك التي تتطلب بالضرورة طرقاً خاصة. وكمثال، يبيّن الجدول (٤ - ١) تصنيف مفردات إلى أربعة أنواع أملتها الخبرة المتوافرة من مسوح إحصائية زراعية تتم على مستوى منطقة، ويعني المسح الإحصائي العام في هذا التصنيف، مسحاً تتوزع فيه الوحدات بصورة عادلة فوق منطقة ما، كما في حالة عيّنة عشوائية بسيطة مثلاً.

جدول (٤ - ١) مثال عن أنواع مختلفة من المفردات في مسوح إحصائية تتم فوق منطقة

النوع	الصفة المميّزة للمفردة	نوع المعاينة الذي نحتاجه
١	انتشار واسع عبر المنطقة وتقع بتكرار معقول في جميع أجزاء المنطقة.	مسح إحصائي عام ولكن بنسبة معاينة منخفضة.
٢	انتشار واسع عبر المنطقة ولكن بتكرار منخفض.	مسح إحصائي عام ولكن بنسبة معاينة أعلى.
٣	تقع بتكرار معقول في معظم أجزاء المنطقة، ولكن بتوزيع متشتت، حيث تكون غائبة في بعض الأجزاء ومركّزة تركيزاً عالياً في أجزاء أخرى.	من أجل أفضل النتائج، نأخذ عيّنة طبقية مع تركيز مختلف في الأجزاء المختلفة من المنطقة (فصل ٥)، ويمكن أحياناً احتواؤها في مسح إحصائي عام مع بعض المعاينات الملحقة.
٤	توزيع متشتت جداً، أو مركّز في جزء صغير من المنطقة.	غير مناسب لمسح إحصائي عام. ويتطلب عيّنة تتكيف بحيث تلائم التوزيع.

(٤ - ٩) حجم العيّنة عندما نريد تقديرات تتعلق بتقسيمات فرعية للمجتمع

غالباً ما تتضمن خطّة المعاينة تقديرات ليس من أجل المجتمع ككل فقط، وإنما من أجل تقسيمات فرعية معينة. وإذا أمكن تحديد هذه التقسيمات سلفاً، كما في حالة مناطق جغرافية مختلفة مثلاً، فإننا نقوم بحساب n في كل منطقة. ولنفرض أننا سنقدّر متوسط كل من الأجزاء الفرعية بتباين محدد V. ففي الجزء i نأخذ $n_i=S_1^2/V$، بحيث يصبح الحجم الكلي للعيّنة $n=\sum S_1^2/V$. وكل S_i^2 بمفرده سيكون، في المتوسط، أصغر من تباين المجتمع S^2، وغالباً ما تكون أصغر بقليل فقط من S^2. وهكذا، إذا

كان هناك k من التقسيمات الفرعية فإن $n \doteq kS^2/V$ بينما سنأخذ $n=S^2/V$ إذا كنا نريد فقط تقديراً يتعلق بالمجتمع ككل.

وهكذا إذا أردنا تقديرات تباينها V في كل من k من التقسيمات الفرعية فسيقترب حجم العيّنة من أن يكون k مرة الحجم n الذي نحتاجه لتقدير يتعلق بالمجتمع ككل وبالدقة نفسها. ويميل الأشخاص غير المتمرسين في طرق المسح الإحصائي إلى إغفال هذه النقطة عند حسابات حجم العيّنة.

وإذا كانت التقسيمات تمثل تصنيفات بوساطة متغيّرات مثل العمر، الجنس، الدخل، وسنيّ الدراسة، فلا يُعرف الجزء الذي ينتمي إليه شخص إلا بعد أخذ العيّنة. ويبقى من الممكن القيام بتقديرات مسبّقة لحجم العيّنة إذا كانت نسب الوحدات المنتمية إلى الأقسام المختلفة π_i، معروفة. وإذا اخترنا عيّنة عشوائية بسيطة حجمها n فإن حجم العيّنة المتوقع من الجزء i هو $n\pi_i$ ومتوسط تباين المتوسط من هذا الجزء هو

$$V(\bar{y}_i) = E\left(\frac{S_i^2}{n_i}\right) \doteq \frac{S_i^2}{n\pi_i} \tag{4.10}$$

إذا كان $n\pi_i$ كبيراً. وهكذا نحتاج إلى $n \doteq S_1^2/\pi_i V$ كي نجعل $V(\bar{y}_i)=V$ وإذا أردنا لهذا أن يكون صحيحاً في كل قسم فعندئذ

$$n \doteq \max\left(\frac{S_i^2}{\pi_i V}\right) \tag{4.11}$$

وإذا كان التقسيم إلى صفوف مثل العمر، الدخل، فيمكن أن يكون S_1^2/π_i أقل من S^2 في صفوف مركزية، إلا أنه يمكن أن يكون أكبر في صف متطرف تصغر فيه π_i. وفي هذه الحالة، إما أن نضطر إلى زيادة قيمة V في هذا القسم أو إيجاد طريقة ما للتعرف على وحدات هذا القسم سلفاً بحيث يمكن أخذ عيّنة منه وفق نسبة أعلى. وأحياناً تكون طريقة المعاينة المضاعفة (فصل ١٢) مفيدة لهذه الغاية.

وتبقى المتطلبات المتعلقة بحجم العيّنة أكبر في الدراسات التحليلية التي نريد أن يتحقق فيها الشرط

$$V(\bar{y}_i - \bar{y}_j) \leq V \tag{4.12}$$

وذلك لكل زوج من التقسيمات الفرعية (الميادين). وفي هذه الحالة

$$n \doteq \max_{i,j} \frac{1}{V}\left(\frac{S_i^2}{\pi_i} + \frac{S_j^2}{\pi_j}\right) \tag{4.13}$$

وإذا كان S_i^2 لا يختلف كثيراً عن S^2 ، فسيكون n مساوياً لِـ $2kS^2/V$ عندما تكون حجوم الميادين الـ k متساوية، وسيبقى أكبر من هذه القيمة فيما عدا ذلك. وتأثير حدود الـ ت م م، التي أهملناها في هذه الدراسة، هو تخفيض الحجم n المطلوب إلى حد ما.

(٤ ـ ١٠) حجم العيّنة في مسائل التقرير

ويمكن أحياناً تطوير أسلوب أكثر منطقياً لتحديد حجم عيّنة وذلك في الحالة التي يكون مطلوباً فيها اتخاذ قرار عملي استنادًا إلى نتائج العيّنة. ويُعتقد أنه في حالة خطأ أقل في تقدير العيّنة فإن القرار سيُبنى على أسس مما لو كان خطأ تقدير العيّنة كبيراً. وقد نستطيع أن نُترجم مالياً الخسارة $l(x)$ التي ستترتب على اتخاذ قرار حيث z مقدار الخطأ في التقدير. ومع أنه لا يمكن التنبؤ سلفاً بالقيمة الفعليّة لِـ z إلا أن نظرية المعاينة تسمح لنا بإيجاد التوزيع التكراري $f(z,n)$ للمتغير z وسيعتمد هذا التوزيع، من أجل طريقة محددة للمعاينة، على حجم العيّنة n . وبالتالي تكون الخسارة المتوقعة مع حجم معطى للعيّنة هي

$$L(n) = \int l(z) f(z, n)\, dz \tag{4.14}$$

والهدف من أخذ العيّنة هو تخفيض هذه الخسارة. وإذا كانت $C(n)$ تكلفة عيّنة حجمها n ، فإن الإجراء المنطقي هو اختيار n الذي يجعل

$$C(n) + L(n) \tag{4.15}$$

أصغر ما يمكن، باعتبار أن هذا يمثل التكلفة الإجمالية لأخذ العيّنة ولاتخاذ قرارات بناء على نتائجها. واختيار n يحدد كلاً من الحجم الأمثل للعيّنة ودرجة الدقة الأفضل. وبصورة بديلة، يمكن عرض الأسلوب نفسه بدلالة الخسارة الناشئة عن أخطاء في معلومات العيّنة. وإذا استخدمنا الربح المالي نضع عبارة للربح المتوقع $G(n)$ من عيّنة حجمها n حيث $G(n)$ يساوي الصفر إذا لم نأخذ عيّنة. ونجعل

$$G(n) - C(n)$$

أعظم ما يمكن، وعلى هذا الشكل يكون المبدأ هنا مكافئاً للقاعدة المتّبعة في الاقتصاد التقليدي وهو جعل الربح أعظم ما يمكن.

وتقع أبسط التطبيقات في الحالة التي تتّخذ فيها دالة الخسارة $l(z)$ الصيغة λz^2 حيث λ عدد ثابت. ونستنتج أن

$$L(n) = \lambda E(z^2) \tag{4.16}$$

وعلى سبيل المثال، إذا كان $\hat{\bar{Y}}$ تقدير العيّنة لِـ $\bar{\bar{Y}}$، و $z = \hat{\bar{Y}} - \bar{Y}$ فإن

$$L(n) = \lambda V(\hat{\bar{Y}}) = \frac{\lambda S^2}{n} - \frac{\lambda S^2}{N} \tag{4.17}$$

في حال استخدام معاينة عشوائية بسيطة.

والشكل الأبسط لدالة التكلفة عند أخذ عيّنة هو

$$C(n) = c_0 + c_1 n \tag{4.18}$$

حيث c_0 التكلفة الابتدائية. وبالاشتقاق نجد أن قيمة n التي تجعل مجموع التكلفة والخسارة أصغر ما يمكن هي

$$n = \sqrt{\lambda S^2 / c_1} \tag{4.19}$$

ويعطي Yates (1960) الشكل الأعم لهذه النتيجة. وينطبق التحليل نفسه على أية طريقة في المعاينة أو التقدير يتناسب تباين التقدير فيها عكساً مع n وتكون التكلفة دالة خطية في n.

ويصف Blythe (1945) تطبيق هذا المبدأ على تقدير حجم الأخشاب في مستودع معدّ للبيع (انـظر التمرين ٤ ـ ١١). ويناقش Nordin (1944) الحجم الأمثل لعيّنة تهدف إلى تقدير المبيعات المحتملة في سوق يعتزم أحد رجال الصناعة دخولها. وإذا أمكن التنبؤ الدقيق بالمبيعات فيمكن تحديد التجهيزات الثابتة وتحديد الإنتاجية للفترة القياسية الواحدة من الإنتاج بحيث تجعل الربح المتوقع لرجل الصناعة أعظم ما يمكن ويدرس Grundy وآخرون (1954, 1956) الحجم الأمثل لعيّنة ثانية بعد معرفتنا لنتائج عيّنة أولى أخذناها.

وقـد تلقّى هذا الأسلوب قدراً كبيراً من التطوير من باحثين في نظرية التقرير الإحصـائية. وتتضمن التعميمات اعتبـار المنفعة بديلاً للقيمة المالية عند وضع سلّم نقيس بموجبه التكاليف والخسائر، كما تتضمن الاستخدام الصريح لمعلومات ذاتية مسبقة تتعلق بمعالم، غير معروفة وذلك بالتعبير عن هذه المعلومات في صيغة توزيعات إحتمالية قَبلية للمعالم المجهولة، وتتضمن تقصّي أنواع مختلفة من دوال الخسارة والتكلفة ودراسـة بيانـات إحصائية نوعية وكميّة على حدّ سواء، وللاطلاع على وصف شامل للطريقة، انظر Raiffa و Schlaifer (1961) ومع أن مدى قدرة هذه الطريقة على تقديم حل كامل لمسألة التقرير لا يزال غير واضح، إلا أن لهذه الطريقة قيمتها في إثارة أفكار واضحة حول العوامل المهمة في مسألة جودة القرار. وأحد الميادين التي تبدو مناسبة للتطبيقات هي معاينة مجموعة ضخمة من المواد في عملية إنتاج كبيرة كي نقرر رفض أو قبول هذه المجموعة على أساس من تقديرنا لنوعيّتها. ويدرس Sittig (1951) المسألة الاقتصادية لتحديد حجم العينة آخذاً في الاعتبار تكاليف التفتيش والتكاليف الناتجة عن العثور على مواد عاطلة في مجموعة مقبولة من البضائع أو العثور على مواد جيدة في مجموعة مرفوضة.

(٤-١١) أثر التصميم (Deff)

ومع خطط المعاينة الأكثر تعقيداً التي سندرسها فيما بعد في هذا الكتاب نجد صفة مفيدة هي صفة أثر التصميم (Deff) الموافق لخطة ما [Kish (1965)]. ويُعرّفها كيش بأنها نسبة تباين التقدير الذي نحصل عليه من العينة (الأكثر تعقيداً) إلى تباين

التقدير الذي نحصل عليه من عيّنة عشوائية بسيطة تتضمن العدد نفسه من الوحدات. ولأثر التصميم استخدامان رئيسيان أحدهما في مجال تقدير العينة والآخر في مجال تثمين كفاءة خطط أكثر تعقيداً. وعلى سبيل المثال، عند تقدير نسبة الناس الذين يمتلكون صفة ما، غالباً ما يكون من المناسب استخدام المنزل كوحدة معاينة بدلاً من الشخص وكما لاحظنا في الفصل الثالث، لا يمكن استخدام الصيغة PQ/n في هذه الخطط. ولتقدير نسبة الذين زارو طبيباً (فقرة ٣-١٢)، أعطت عينة عشوائية بسيطة من المنازل $V(p)=0.00520$ مقابل $pq/n=0.00197$ في عيّنة عشوائية بسيطة من الحجم نفسه ولكنها عينة من الأشخاص. وتقدير أثر التصميم لهذه المعاينة العنقودية هو 520/197=2.6 . وعندما تكون كسور المعاينة صغيرة يمكننا تقدير حجم عينة بحساب الـ n (عدد الأشخاص) الذين نحتاجهم في عينة عشوائية بسيطة من الأشخاص وضربهِ بـ 2.6 ، وبهذه الطريقة إذا لاحظنا أثر التصميم لمتغيرات مهمة في خطة معقدة، فيمكن استخدام الصيغ البسيطة في هذا الفصل لتقدير حجم العينة في الخطة المعقدة والحكم على ما إذا كانت الخطة المعقدة مفيدة من حيث كفاءتها وذلك بالمقارنة مع تكلفتها وتعقيدها. وقد نحتاج إلى بعض الجبر عند تقدير أثر التصميم من نتائج عينة معقدة. ونحتاج إلى تبيان الكيفية التي تقدم هذه النتائج بموجبها، إذا أمكن، تقديرات غير منحازة للتباين ولِـ S^2 . وقد أعطيت أمثلة من هذه الحسابات في حالة معاينة عشوائية طبقية في الفقرة (٥ ـ ١١ ا)، وفي حالة معاينة عنقودية بعناقيد متساوية الحجم في الفقرة (٩ ـ ٣).

تمارين

(٤ ـ ١) في منطقة تتضمن 4000 منزل نريد تقدير نسبة البيوت التي يقطنها مالكوها بخطأ معياري لا يزيد على 2% ، ونسبة الأسر التي تمتلك سيارتين بخطأ معياري لا يزيد على 1% . (الرقمان 2% و 1% هما القيمتان المطلقتان، وليسا معامل الاختلاف) ويظن أن نسبة المالكين تقع بين 45% و 65% وأن نسبة الأسر التي تمتلك سيارتين تقع بين 5% و 10% . ما هو حجم العيّنة الضروري لتحقيق الهدفين؟

(٤ - ٢) في مجتمع مؤلف من 676 صفحة من صفحات معروض (جدول ٢-٢ صفحة. . .) كم يجب أن يكون حجم العينة إذا كنا سنقدر العدد الكلي للتواقيع بهامش خطأ يساوي 1000 ، وذلك باستثناء فرصة واحدة من عشرين؟ افترض أن S^2 المعطى في الصفحة رقم ٤١ هو S^2 الخاص بالمجتمع.

(٤ - ٣) نقوم بمسح إحصائي لتفشّي الأمراض العامة في مجتمع كبير، ونرغب لكل مرض يصيب 1 بالمائة، على الأقل، من أفراد المجتمع، تقدير العدد الكلي للإصابات بمعامل اختلاف لا يزيد على 20 في المائة.

(i) ما هو حجم العيّنة العشوائية البسيطة التي نحتاجها، بفرض أنه يمكن التعرف على وجود المرض (تشخيصه) بدون أخطاء؟

(ii) ماهو الحجم الذي نحتاجه إذا كنا نريد عدد الحالات الكلي لكل من الذكور والإناث على حدة، وبالدقة ذاتها؟

(٤ - ٤) في مسح إحصائي للدود الشريطي، كان علينا تقدير عدد الدود الشريطي في الفدان بحدود للخطأ تساوي 30 بالمائة، وبمعامل ثقة يساوي 95 بالمائة، وذلك في كل حقل تتجاوز كثافة الدود الشريطي في تربته السطحية، وحتى عمق 5 بوصات، الـ 200,000 في الفدان. وتقيس أداة المعاينة كتلة ترابية قياسها 5 x 9 x 9 بوصة. وبفرض أن عدد الدود الشريطي في عينة واحدة يتبع توزيعاً أكثر تغيراً بقليل من توزيع بواسون، نأخذ $S^2=1.2\bar{Y}$. ما هو حجم العيّنة العشوائية البسيطة التي نحتاجها؟ (الفدان يساوي 43560 قدماً مربعاً).

(٤ - ٥) في مسح إحصائي يتناول المزارع في ولاية أيوا، حصلنا على معاملات الاختلاف التالية على أساس الوحدة، علماً بأن الوحدة هي مساحة ميل مربع (البيان الإحصائي لـ R. J. Jessen).

تقدير معامل الاختلاف %	المفردة
38	فدادين معدّة للزراعة
39	فدادين قمح
44	فدادين شوفان
100	عدد العمال في الأسرة
110	عدد العمال المستأجرين
317	عدد العاطلين عن العمل

وقد خُطّط مسح إحصائي لتقدير المفردات المتعلقة بالفدادين بمعامل اختلاف يساوي $2\frac{1}{2}$ في المائة ولتقدير أعداد العمال (باستثناء غير المستخدمين منهم) بمعامل اختلاف يساوي 5 بالمائة. فما هو عدد الوحدات التي نحتاجها في حالة معاينة عشوائية بسيطة؟ ما هو مدى الجودة المتوقعة في مقدرة هذه العيّنة على تقدير عدد غير المستخدمين؟

(٤ ـ ٦) من خلال معاينة تجريبية، نريد تقدير القيمة المتوسطة لمتغير عشوائي وذلك بتباين $V = 0.0005$ للعيّنات العشرين الأولى المسحوبة، فكم هو عدد العينات الإضافية التي نحتاجها؟ [استخدم المعادلة (4.7)].

قيمة المتغير العشوائي	رقم العينة	قيمة المتغير العشوائي	رقم العينة
0.0712	11	0.0725	1
0.0748	12	0.0755	2
0.0878	13	0.0759	3
0.0710	14	0.0739	4
0.0754	15	0.0732	5
0.0712	16	0.0843	6
0.0757	17	0.0727	7
0.0737	18	0.0769	8
0.0704	19	0.0730	9
0.0723	20	0.0727	10

(٤ - ٧) صُمّم مسح منزلي لتقدير نسبة الأسَر التي تمتلك صفات معينة. ومن أجل المفردات الأكثر أهمية نتوقع أن تقع قيمة P بين 30% و 70% ما هي، في معاينة عشوائية بسيطة، قيم n الضرورية لتقدير المتوسطات التالية بخطأ معياري لا يتجاوز 3%؟

(أ) المتوسط الإجمالي P ؟

(ب) المتوسطات الإفرادية P_i لفئات الدخل - تحت 5000$ من 5000$ إلى 10000$ فوق الـ 10000$ $(i=1,2,3)$ ؟

(ج) الفروق بين المتوسطات (P_i-P_j) لكل زوج من الفئات في (ب) ؟ أعطِ جواباً منفصلاً لكل من (أ)، (ب)، (ج). وتشير إحصاءات الدخل إلى أن نسب العائلات ذات الدخل في الفئات الثلاث المذكورة أعلاه هي 50% ، 38% و 12% .

(٤ - ٨) قُسمت الكليات ذات الأربع سنوات في الولايات المتحدة إلى فئات من أربعة حجوم مختلفة وذلك وفقاً لرقم تسجيلها عام 1952-1953 والانحرافات المعيارية ضمن كل فئة مبيّنة أدناه.

	الفئة			
	1	2	3	4
عدد الطلاب	<1000	1000–3000	3000–10,000	over 10,000
S_i	236	625	2008	10,023

إذا علمت حدود الفئات دون قيم S_i ، ما هو مدى الجودة التي يمكنك فيها تخمين قيم S_i مستخدماً أرقاماً رياضية بسيطة (فقرة ٤-٧) ؟ لاتوجد كلية بأقل من 200 طالب، وأكبرها يحوي حوالي 50,000 طالب.

(٤-٩) مع دالة خسارة تربيعية ودالة تكلفة خطية، كما في الفقرة ٤ - ١٠، انخفضت S^2 إلى S'^2 بوساطة خطة معاينة متفوقة، مع بقاء c_0 ، c_1 . و λ بدون تغير. إذا رمزنا بـ n' ، V' لحجم العيّنة الأمثل الجديد ولـ $V(\hat{y})$ المرافق، برهن أن $n'<n$ وأن $V'<V$.

(٤ ـ ١٠) إذا كانت دالة الخسارة التي يسببها خطأ في $\bar{y}$ هو $\lambda|\bar{y}-\bar{Y}|$ ، وكانت التكلفة $C=c_0+c_1n$ فبيّن أن القيمة الأكثر اقتصادية لِـ n ، في حالة معاينة عشوائية بسيطة، ومتجاهلين عامل الـ ت م م هي

$$\left(\frac{\lambda S}{c_1\sqrt{2\pi}}\right)^{2/3}$$

(٤ ـ ١١) (مقتبسة من Blythe, 1945) ثمن مبيع قطعة أرض من أشجار الغابات الصالحة لصناعة الأخشَاب هو UW حيث U هو سعر وحدة الحجوم و W حجم الخشب في هذه القطعة. أحصينا عدد الجذوع N في هذه القطعة، ثم قدرنا، من عينة عشوائية بسيطة من n من هذه الجذوع، الحجم الوسطي للجذع الواحد. ويقوم البائع بهذا التقدير كما يدفع تكلفته، ويوافق عليه المشتري مؤقتاً. وفيما بعد يكتشف المشتري حجم ما تمّ شراؤه بالضبط، ويعوض عليه البائع إذا كان قد دفع لأكثر مما تم تسليمه. وإذا كان ما دفعه أقل مما تستحقه البضاعة المسلّمة فإن المشتري لا يذكر الحقيقة.

اكتب دالّة خسارة البائع. وبفرض أن تكلفة قياس n من الجذوع هي cn ، فاحسب القيمة المثلى لِـ n . ويمكن أن نرمز بِـ S للانحراف المعياري للحجم على أساس الجذع الواحد، كما يمكن تجاهل عامل الـ ت م م.

(٤ ـ ١٢) (ا) نريد معرفة حضور أو غياب كل من صفتين في كل وحدة من وحدات عينة عشوائية بسيطة مأخوذة من مجتمع كبير. إذا كانت P_1 و P_2 النسبتين المئويتين لوحدات المجتمع التي تملك الصفتين 1 و 2 ، على الترتيب، ويرغب زبون في تقدير (P_1-P_2) بخطأ معياري لا يتجاوز الاثنين في المائة. فما حجم العيّنة الذي تقترحه إذا كان الزبون يعتقد أن P_1 و P_2 تقعان بين 40% و 60% وأن الصفتين تتوزعان بصورة مستقلة على الوحدات؟
(ب) لنفرض أن الزبون في (ا) يعتقد أن الصفتين مرتبطتان إيجاباً، ولكنه لا يعلم قيمة معامل الارتباط. وأنك اقترحت عيّنة ابتدائية حجمها 200 تمخضت عن النتائج التالية:

عدد الوحدات	الصفة 1	الصفة 2
72	نعم	نعم
44	لا	نعم
14	نعم	لا
70	لا	لا
200		

فما هو حجم العيّنة الذي تقترحه الآن لتقدير (P_2-P_1) بخطأ معياري لا يتجاوز 2%؟

(٤ ـ ١٣) (ا) لنفرض أنك تقدر نسبة الجنس، وهي قريبة من النصف، ويمكنك معاينة منازل تضم أربعة أشخاص: الأب، الأم، وطفلين. متجاهلاً النسب الصغيرة من الأسر التي تتضمن توأماً متطابقاً، أوجد أثر التصميم في حالة عينة عشوائية بسيطة تتضمن n منزلاً في مقابل عيّنة من الـ $4n$ شخصاً.

(ب) هل تخفض العائلات التي تمتلك التوأم المتطابق أثر التصميم أم ترفعه؟

المعاينة العشوائيّة الطبقيّة

(٥ - ١) مقدمـــة

أول ما نقوم به في المعاينة الطبقية هو تقسيم المجتمع المؤلف من N وحدة إلى مجتمعات جزئية فيهـا $N_1, N_2, ..., N_L$ من الـوحدات على الترتيب. وهذه المجتمعات الجزئية غير متداخلة، وهي تؤلف مع بعضها المجتمع بكامله، أي أن:

$$N_1 + N_2 + ... N_L = N$$

وتسمى المجتمعات الجزئية طبقات. وللحصول على الفائدة التامة من عملية التقسيم إلى طبقات يجب معرفة قيم المقادير N_h. وعند تحديد الطبقات، تسحب عيّنة من كل طبقـة، ويتم السحب بصـورة مستقلة في الطبقات المختلفة. ونرمز لحجوم العيّنات ضمن الطبقات بـ $n_1, n_2, ..., n_L$ على الترتيب.

وإذا أخذنا عيّنة عشوائية بسيطة من كل طبقة، توصف عندئذ مجمل الطريقة بأنها معاينة عشوائية طبقية.

والتقسيم الى طبقات طريقة عامة جداً، وهناك أسباب كثيرة لذلك أهمها ما يلي:

(١) إذا أردنا معلومات إحصائية، وبدقة معروفة، لأجزاء معينة من المجتمع، فمن المستحسن أن نعالج كل جزء وكأنه «مجتمع» قائم بذاته.

(٢) وقد تملي الراحة في العمل الإداري استخدام التقسيم إلى طبقات، فمثلاً قد يكون للوكالة التي تقوم بمسح إحصائي دوائر ميدانية، تشرف كل دائرة منها على المسح المتعلق بجزء من المجتمع.

(٣) قد تختلف مشاكل المعاينة بصورة ملحوظة في أجزاء مختلفة من المجتمع. وفي المجتمعات البشرية، غالباً ما يوضع الناس الذين يعيشون في مؤسسات (مثلاً:

فنادق، مستشفيات، سجون) في طبقة مختلفة عن أولئك الذين يعيشون في بيوت عادية، لأن طرق المعاينة المناسبة للحالتين مختلفة. وقد نمتلك في مسائل المعاينة قائمة بالشركات الكبرى التي نضعها في طبقة منفصلة. وقد نضطر لاستخدام نوع من المعاينة التي تستخدم المساحات كوحدات معاينة في الشركات الأصغر.

(٤) يمكن أن يؤدي التقسيم إلى طبقات إلى كسب في دقة تقديرات صفات مميزة للمجتمع ككل. والفكرة الأساسية هي أنه قد يكون من الممكن تقسيم مجتمع غير متجانس إلى مجتمعات جزئية يتصف كل منها بأنه متجانس داخلياً. وهذا ما يوحي به اسم «الطبقات» بكل ما تتضمنه من معنى التقسيم إلى أجزاء. وإذا كانت كل من الطبقات متجانسة، بمعنى أن تختلف القياسات قليلاً جداً من وحدة إلى أخرى، فيمكن الحصول على تقدير دقيق لمتوسط أي طبقة من خلال عيّنة صغيرة ضمن هذه الطبقة. ونستطيع، عندئذ تركيب هذه التقديرات في تقدير واحد دقيق يخص المجتمع ككل.

وتعالج نظرية المعاينة الطبقية خواص التقديرات من عيّنة طبقية كما تعالج مسألة أفضل اختيار لحجوم العيّنات بحيث نحصل على أعظم دقة ممكنة. ونفترض في هذه المرحلة من المناقشة أن الطبقات قد تمّ تشكيلها. والمسائل المتعلقة بكيفية بناء الطبقات وعدد الطبقات اللازمة، ستؤجّل إلى مرحلة متأخرة [فقرة (٥-٧)].

(٥ - ٢) رمــوز

يرمز الدليل h للطبقة و i للوحدة ضمن الطبقة. والرموز تعميم طبيعي لتلك المستخدمة سابقاً. وتشير جميع الرموز التالية إلى طبقة h

العدد الكلي للوحدات N_h

عدد الوحدات في العيّنة n_h

القيمة التي نحصل عليها من أجل الوحدة i y_{hi}

ترجيحة الطبقة

$$W_h = \frac{N_h}{N}$$

كسر المعاينة ضمن الطبقة $f_h = \frac{n_h}{N_h}$

المتوسط الصحيح $\bar{Y}_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h}$

متوسط العيّنة $\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$

التباين الصحيح $S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2}{N_h - 1}$

لاحظ أن المقام في علاقة التباين هو (N_h-1)

(٥ - ٣) خـواص التقديـرات

من أجـل متـوسط المجتمـع على أسـاس الـوحدة الواحدة، نجد أن التقدير المستخدم في المعاينة الطبقية هو $\bar{y}_{st}$ مختصر Stratified حيث:

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N} = \sum_{h=1}^{L} W_h \bar{y}_h \qquad (5.1)$$

وحيث $N = N_1 + N_2 + \ldots + N_L$

ولا يكون التقدير $\bar{y}_{st}$ بصورة عامة، هو متوسط العيّنة نفسه، ذلك لأنه يمكن كتابة متوسط العيّنة $\bar{y}$ على الشكل

$$\bar{y} = \frac{\sum_{h=1}^{L} n_h \bar{y}_h}{n} \qquad (5.2)$$

والفرق هو أنـه في $\bar{y}_{st}$ تتلقى التقـديرات من الـطبقـات كل بمفـردهـا ترجيحـاتها الصحيحة N_h/N . ومن الواضح أن تتطابق مع $\bar{y}_{st}$ شريطة أن يتحقق في كل طبقة

$$\frac{n_h}{n} = \frac{N_h}{N} \quad \text{أو} \quad \frac{n_h}{N_h} = \frac{n}{N} \quad \text{أو} \quad f_h = f$$

وهذا يعني أن كسر المعاينة يبقى نفسه في جميع الطبقات. ويوصف مثل هذا التقسيم إلى طبقات بأنه تقسيم إلى طبقات بحصص متناسبة مع n_h أو المحاصّة التناسبية. وهو يعطي عينة ذاتية الترجيح. وإذا كان لدينا العديد من التقديرات لنقوم بها فإن العينة ذاتية الترجيح توفر الوقت.

ونُجمل في النظريات التالية الخواص الرئيسة للتقدير $\bar{y}_{st}$. وتنطبق النظريتان الأولى والثانية على المعاينة الطبقية، بصورة عامة، وهذا يعني أنه ليس من الضروري أن تكون العيّنة المأخوذة من كل طبقة عيّنة عشوائية بسيطة.

نظرية (٥ - ١)

إذا كان تقدير العيّنة $\bar{y}_h$ غير منحاز في كل طبقة، فيكون $\bar{y}_{st}$ عندئذ تقديراً غير منحاز لمتوسط المجتمع $\bar{Y}$.

برهان

$$E(\bar{y}_{st}) = E\sum_{h=1}^{L} W_h \bar{y}_h = \sum_{h=1}^{L} W_h \bar{Y}_h$$

باعتبار أن التقديرات في كل طبقة هي تقديرات غير منحازة. ولكن يمكن كتابة متوسط المجتمع $\bar{Y}$ على الشكل،

$$\bar{Y} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{N_h} y_{hi}}{N} = \frac{\sum_{h=1}^{L} N_h \bar{Y}_h}{N} = \sum_{h=1}^{L} W_h \bar{Y}_h$$

وهو المطلوب.

نظرية (٥ - ٢)

إذا سحبنا بصورة مستقلة عيّنتين من طبقتين مختلفتين فعندئذ،

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 V(\bar{y}_h) \tag{5.3}$$

حيث $V(\bar{y}_h)$ هو تباين $\bar{y}_h$ فوق عيّنات متكررة من الطبقة h .

برهان

بما أن

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{5.4}$$

فنجد أن $\bar{y}_{st}$ دالة خطية في المقادير بترجيحات ثابتة W_h وبالتالي يمكن اقتباس النتيجة الإحصائية المتعلقة بتباين دالة خطية .

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} W_h^2 V(\bar{y}_h) + 2 \sum_{h=1}^{L} \sum_{j>h}^{L} W_h W_j \operatorname{Cov}(\bar{y}_h \bar{y}_j) \tag{5.5}$$

وبما أن العيّنتين مسحوبتان بصورة مستقلة من طبقتين مختلفتين فجميع حدود التغاير تنعدم ، وهذا يؤدي إلى النتيجة (5.3) .

ونلخص النظريتين (٥ ـ ١) و(٥ ـ ٢) كما يلي : إذا كان $\bar{y}_h$ تقديراً غير منحاز لـ $\bar{Y}_h$ في كل طبقة ، وكان اختيار العيّنات مستقلاً في الطبقات المختلفة ، فعندئذ يكون $\bar{y}_{st}$ تقديراً غير منحاز لـ $\bar{Y}$ بتباين يساوي $\sum W_h^2 V(\bar{y}_h)$ والنقطة المهمة في هذه النتيجة هي أن تباين $\bar{y}_{st}$ يعتمد فقط على تباينات تقديرات المتوسطات $\bar{Y}_h$ للطبقات كل بمفردها .

وإذا أمكن تقسيم مجتمع شديد التغير إلى طبقات بحيث يكون لجميع المفردات القيمة نفسها ضمن طبقة واحدة ، فيمكن عندئذ تقدير $\bar{Y}$ بدون أي خطأ . ويبيّن تأملنا للبرهان أن استخدام الأوزان الصحيحة N_h/N في كل طبقة هو الذي قادنا إلى هذه النتيجة .

نظرية (٥ ـ ٣)

في معاينة عشوائية طبقية ، يكون تباين التقدير

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h} = \sum_{h=1}^{L} W_h^2 \frac{S_h^2}{n_h} (1 - f_h) \tag{5.6}$$

برهان

بما أن تقدير $\bar{y}_h$ غير منحاز لـ $\bar{Y}_h$ فيمكن تطبيق النظرية (٥ ـ ٢). ومن النظرية (٢ ـ ٢) مطبقة على طبقة بمفردها نجد:

$$V(\bar{y}_h) = \frac{S_h^2}{n_h} \frac{N_h - n_h}{N_h}$$

وبالتعويض في نتيجة النظرية (٥ ـ ٢) نحصل على

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h^2 V(\bar{y}_h) = \frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h} = \sum W_h^2 \frac{S_h^2}{n_h} (1 - f_h)$$

ونقدم في النتائج التالية بعض الحالات الخاصة لهذه العلاقة

نتيجة (١)

إذا كان كسر المعاينة n_h/N_h مهملاً في جميع الطبقات فإن:

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum \frac{N_h^2 S_h^2}{n_h} = \sum \frac{W_h^2 S_h^2}{n_h} \qquad (5.7)$$

وهي العلاقة المناسبة عندما يكون إهمال عامل الـ (ت م م) ممكناً.

نتيجة (٢)

في حالة الحصص المتناسبة، نعوض

$$n_h = \frac{n N_h}{N}$$

في (5.6) . فيصبح التباين على الشكل

$$V(\bar{y}_{st}) = \sum \frac{N_h}{N} \frac{S_h^2}{n} \left(\frac{N - n}{N} \right) = \frac{1 - f}{n} \sum W_h S_h^2 \qquad (5.8)$$

نتيجة (٣)

إذا كانت المعاينة تناسبية وكان للتباينات في جميع الطبقات القيمة S_w^2 نفسها، فنحصل على النتيجة البسيطة التالية

$$V(\bar{y}_{st}) = \frac{S_w^2}{n}\left(\frac{N-n}{N}\right) \tag{5.9}$$

نظرية (٥ ـ ٤)

إذا كان $\hat{Y}_{st} = N\bar{y}_{st}$ هو تقدير لمجموع المجتمع Y فعندئذ

$$V(\hat{Y}_{st}) = \sum N_h(N_h - n_h)\frac{S_h^2}{n_h} \tag{5.10}$$

وهذا ينتج مباشرة من النظرية (٥ ـ ٣).

مثال

يبين الجدول (٥ ـ ١) تعدادي 1920 و 1930 لسكان 64 من المدن الكبرى في الولايات المتحدة، بالآلاف. وقد حصلنا على البيان الإحصائي بأخذ المدن التي يأتي ترتيبها بين الخامسة والثامنة والستين في الولايات المتحدة وفقاً لعدد سكانها في عام 1920. وقد رتبنا المدن في طبقتين، الأولى تحوي المدن الست عشرة الأكبر وتحوي الثانية المدن الثماني والأربعين الباقية.

جدول (٥ ـ ١) حجوم 64 مدينة (بالآلاف) في 1920 و 1930

حجم عام 1920 (x_{hi})				حجم عام 1930 (y_{hi})			
الطبقة				الطبقة			
$h = 1$	2			1	2		
797	314	172	121	900	364	209	113
773	298	172	120	822	317	183	115
748	296	163	119	781	328	163	123
734	258	162	118	805	302	253	154
588	256	161	118	670	288	232	140
577	243	159	116	1238	291	260	119
507	238	153	116	573	253	201	130
507	237	144	113	634	291	147	127
457	235	138	113	578	308	292	100
438	235	138	110	487	272	164	107
415	216	138	110	442	284	143	114
401	208	138	108	451	255	169	111
387	201	136	106	459	270	139	163
381	192	132	104	464	214	170	116
324	180	130	101	400	195	150	122
315	179	126	100	366	260	143	134

ملاحظة

المدن مذكورة بالترتيب نفسه في كل من العامين.

المجاميع ومجاميع المربعات

طبقة	1920		1930	
	$\sum(x_{hi})$	$\sum(x_{hi}^2)$	$\sum(y_{hi})$	$\sum(y_{hi}^2)$
1	8,349	4,756,619	10,070	7,145,450
2	7,941	1,474,871	9,498	2,141,720

وسنقدّر تعداد السكان عام 1930 في جميع المدن الأربع والستين من عينة حجمها 24 أوجد الخطأ المعياري للمجموع المقدّر في حالة (١) عيّنة عشوائية بسيطة. (٢) عيّنة عشوائية طبقية بحصص متناسبة، (٣) عيّنة عشوائية طبقية باثنتي عشرة وحدة من كل طبقة.

وهذا المجتمع شبيه بمجتمعات أنواع عديدة من المشروعات التجارية من حيث إن بعض الوحدات (المدن الكبرى) تسهم بشكل أكبر بكثير في المجموع الكلي وتُظهر قدراً من التغير أكبر بكثير من الوحدات الباقية.

ويعطي الجدول الملحق بالجدول (٥ ـ ١) مجاميع الصفات ومجاميع المربعات. ونستخدم في هذا المثال بيانات 1930 فقط. وستظهر بيانات 1920 في مثال لاحق.

وفيما يتعلق بعدد السكان عام 1930 نجد

$$Y = 19,568, \qquad S^2 = 52,448$$

ونرمز للتقديرات الثلاثة لِـ Y بِـ $\hat{Y}_{ran}$ ، $\hat{Y}_{prop}$ و $\hat{Y}_{equal}$ ،

(١) من أجل المعاينة العشوائية البسيطة

$$V(\hat{Y}_{ran}) = \frac{N^2S^2}{n}\frac{N-n}{N} = \frac{(64)^2(52,448)}{24}\left(\frac{40}{64}\right) = 5,594,453$$

ومن النظرية (٢-٢) نتيجة ٢ نجد أن الخطأ المعياري

$$\sigma(\hat{Y}_{ran}) = 2365$$

(٢) في كل من الطبقتين، نجد أن التباين:

$$S_1^2 = 53{,}843, \qquad S_2^2 = 5581$$

ونلاحظ أن التباين بين المدن الأكبر يساوي تقريباً 10 أمثال التباين في الطبقة الأخرى. ولدينا في الحصص المتناسبة، $n_1=6$ ، $n_2=18$ ومن العلاقة (5.7) بعد الضرب بـ N^2 نجد

$$V(\hat{Y}_{prop}) = \frac{N-n}{n}\sum N_h S_h^2$$
$$= \tfrac{40}{24}[(16)(53{,}843) + (48)(5581)] = 1{,}882{,}293$$
$$\sigma(\hat{Y}_{prop}) = 1372$$

(٣) من أجل $n_1=n_2=12$ نستخدم العلاقة العامة (5.7) فنجد

$$V(\hat{Y}_{equal}) = \sum N_h(N_h-n_h)\frac{S_h^2}{n_h}$$
$$= \frac{(16)(4)(53{,}843)}{12} + \frac{(48)(36)(5581)}{12} = 1{,}090{,}827$$
$$\sigma(\hat{Y}_{equal}) = 1044$$

ونجد في هذا المثال أن الحجوم المتساوية للعينتين من الطبقتين أكثر دقة من الحصص المتناسبة. وكلاهما أفضل بكثير من المعاينة العشوائية البسيطة.

(٥ ـ ٤) تقدير التباين وحدود الثقة

إذا أخذنا عيّنة عشوائية بسيطة ضمن كل طبقة، فإن التقدير غير المنحاز للتباين S_h^2 هو (وفقاً للنظرية ٢ ـ ٤)

$$s_h^2 = \frac{1}{n_h-1}\sum_{i=1}^{n_h}(y_{hi}-\bar{y}_h)^2 \tag{5.11}$$

ومنه نجد النظرية التالية:

نظرية (٥ ـ ٥)

في المعاينة العشوائية الطبقية يكون التقدير التالي تقديراً غير منحاز لتباين $\bar{y}_{st}$

$$v(\bar{y}_{st}) = s^2(\bar{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{L} N_h(N_h-n_h)\frac{s_h^2}{n_h} \tag{5.12}$$

ويمكن كتابة هذه العبارة بشكل آخر مناسب للأعمال الحسابية

$$s^2(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 s_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h s_h^2}{N} \qquad (5.13)$$

ويمثل الحد الثاني في الطرف الأيمن التخفيض العائد لعامل الـ ت م م.

ولحساب هذا التقدير، يجب أن يكون لدينا وحدتان، على الأقل، مسحوبتان من كل طبقة. ونناقش في الفقرة [(٥ ـ ١٢) ا] تقدير التباين عندما نمضي في التقسيم إلى طبقات إلى الحد الذي نختار فيه وحدة واحدة من كل طبقة. والعلاقات الخاصة بحدود الثقة هي كما يلي:

متوسط المجتمع $$\bar{y}_{st} \pm ts(\bar{y}_{st}) \qquad (5.14)$$

مجموع المجتمع $$N\bar{y}_{st} \pm tNs(\bar{y}_{st}) \qquad (5.15)$$

وتفترض هاتان العلاقتان أن $\bar{y}_{st}$ يتوزع طبيعياً، وأن $s(\bar{y}_{st})$ محدد تحديداً جيداً، بحيث يمكن قراءة العامل t من جداول التوزيع الطبيعي.

وإذا قدّمت كل طبقة عدداً قليلاً من درجات الحرية، فإن الطريقة المعتادة لحساب خطأ العيّنة الموافق لكمية مثل $s(\bar{y}_{st})$ هي أن نقرأ القيمة t من جداول توزيع ستيودنت بدلاً من جدول التوزيع الطبيعي. وبصورة عامة يكون توزيع $s(\bar{y}_{st})$ من التعقيد بحيث لا يسمح بتطبيق دقيق لهذه الطريقة. والطريقة التقريبية لتخصيص عدد فعّال من درجات الحرية لـ $s(\bar{y}_{st})$ هي كما يلي [1946, Satterthwaite].

يمكننا كتابة:

$$s^2(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} g_h s_h^2, \quad \text{حيث} \quad g_h = \frac{N_h(N_h - n_h)}{n_h}$$

والعدد الفعّال من درجات الحرية n_e هو:

$$n_e = \frac{\left(\sum g_h s_h^2\right)^2}{\sum \frac{g_h^2 s_h^4}{n_h - 1}} \qquad (5.16)$$

وتقع قيمة n_e دائماً بين أصغر قيم الكميات (n_h-1) وبين مجموع هذه الكميات. ويأخذ التقريب في الاعتبار حقيقية أن S_h^2 يمكن أن يتغير من طبقة إلى طبقة. ونحتاج هنا إلى الفرض بأن المتغيرات y_{hi} تتوزع وفق التوزيع الطبيعي، باعتبار أن التقريب يعتمد على نتيجة أن تباين s_h^2 هو $2\sigma_h^4/(n_h-1)$. وإذا كان لتوزيع y_{hi} تفرطح إيجابي فسيكون تباين s_h^2 أكبر من ذلك، وستبالغ العلاقة (5.16) في تقدير العدد الفعّال من درجات الحرية.

(٥ - ٥) المحاصّة المثلى

في المعاينة الطبقية يختار المعاين قيم حجوم العيّنة n_h في الطبقات المتتالية. وقد يختارها بحيث تجعل $V(\bar{y}_{st})$ أصغر ما يمكن من أجل تكلفة محددة للحصول على العيّنة، أو بحيث تجعل التكلفة أصغر ما يمكن من أجل قيمة محددة لـ $V(\bar{y}_{st})$ وأبسط دالة تكلفة هي من الشكل

$$\text{التكلفة} = C = c_0 + \sum c_h n_h \tag{5.17}$$

أي تكون التكلفة ضمن كل طبقة متناسبة مع حجم العيّنة، إلا أن تكلفة وحدة المعاينة c_h يمكن أن تتغير من طبقة إلى طبقة. ويمثّل الحد c_0 التكلفة الابتدائية. وتكون دالة التكلفة هذه مناسبة عندما يكون الشيء الرئيس في التكلفة هو أخذ القياسات في كل وحدة معاينة. وإذا كانت تكاليف الانتقال بين الوحدات كبيرة فإن الدراسات الرياضية والتجريبية تقترح أن أفضل تمثيل لتكاليف الانتقال يكون بوساطة العبارة $\sum t_h\sqrt{n_h}$ حيث t_h معدّل تكلفة الانتقال للوحدة الواحدة [Beardwood وآخرون (1959)] ونعتبر هنا دالة التكلفة الخطية (5.17) فقط.

نظرية (٥ - ٦)

في معاينة عشوائية طبقية مع دالة تكلفة خطيّة من الشكل (5.17) يكون تباين تقدير المتوسط أصغر ما يمكن من أجل تكلفة محددة C وتكون التكلفة C أصغر ما يمكن من أجل تباين محدد $V(\bar{y}_{st})$ عندما يتناسب n_h مع $W_hS_h/\sqrt{c_h}$.

برهان

لدينا

$$C = c_0 + \sum_{h=1}^{L} c_h n_h \qquad (5.17)$$

$$V = V(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h}(1-f_h) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{N_h} \qquad (5.18)$$

ومسألتانا هما (١) إما اختيار الـ n_h بحيث يكون V أصغر ما يمكن من أجل C محددة، أو (٢) اختيار الـ n_h بحيث تكون C أصغر ما يمكن من أجل V محدد. ويتفق أن يكون للمسألتين الحل نفسه إذا استثنينا الخطوات الأخيرة. فاختيار المقادير n_h بحيث نجعل V أصغر ما يمكن من أجل C مثبتة أو جعل C أصغر ما يمكن من أجل V مثبت، يكافىء جعل الجداء

$$V'C' = \left(V + \sum \frac{W_h^2 S_h^2}{N_h}\right)(C - c_o)$$
$$= \left(\sum \frac{W_h^2 S_h^2}{n_h}\right)\left(\sum c_h n_h\right) \qquad (5.19)$$

أصغر ما يمكن.

وقد لاحظ Stuart (1945) أنه يمكن بسهولة جعل (5.19) أصغر ما يمكن وذلك باستخدام متراجحة كوشي ـ شوارتز. وإذا كانت a_h ، b_h مجموعتين من L من الأعداد الموجبة، فتأتي هذه المتراجحة من المطابقة

$$\left(\sum a_h^2\right)\left(\sum b_h^2\right) - \left(\sum a_h b_h\right)^2 = \sum_i \sum_{j>i} (a_i b_j - a_j b_i)^2 \qquad (5.20)$$

ونستنتج من (5.20) أن

$$\left(\sum a_h^2\right)\left(\sum b_h^2\right) \geq \left(\sum a_h b_h\right)^2 \qquad (5.21)$$

وتتحقق المساواة إذا وفقط إذا كانت b_h / a_h ثابتة من أجل جميع قيم h وفي (5.19) إذا أخذنا

$$a_h = \frac{W_h S_h}{\sqrt{n_h}}, \qquad b_h = \sqrt{c_h n_h}, \qquad a_h b_h = W_h S_h \sqrt{c_h}$$

فإن المتراجحة (5.21) تعطي

$$V'C = \left(\sum \frac{W_h^2 S_h^2}{n_h}\right)\left(\sum c_h n_h\right) = \left(\sum a_h^2\right)\left(\sum b_h^2\right) \geq \left(\sum W_h S_h \sqrt{c_h}\right)^2$$

وهكـذا فإنـه لا يوجـد اختيار للأعـداد n_h يمكن أن يجعـل $V'C'$ أصغـر من $(\sum w_h S_h \sqrt{C_h})$. وتقع النهاية الصغرى عندما يكون

$$\frac{b_h}{a_h} = \frac{n_h \sqrt{c_h}}{W_h S_h} = \text{ثابت} \tag{5.22}$$

كما ورد في نص النظرية.

وبدلالة حجم العيّنة الكلي n_h في طبقة، نجد

$$\frac{n_h}{n} = \frac{W_h S_h / \sqrt{c_h}}{\sum (W_h S_h / \sqrt{c_h})} = \frac{N_h S_h / \sqrt{c_h}}{\sum (N_h S_h / \sqrt{c_h})} \tag{5.23}$$

وتقود هذه النظرية إلى القواعد الإجرائية التالية. في طبقة معيّنة، خذ عيّنة أكبر إذا كانت:

١ ـ الطبقة أكبر،

٢ ـ التغيرات الداخلية في الطبقة أكبر،

٣ ـ المعاينة من الطبقة أرخص.

ونحتـاج إلى خطوة إضـافية لإتمـام المحـاصّة. إذ تعطي المعادلة (5.23) قيمة n_h بدلالة n ، ولكننا لا نعرف بعدُ قيمة n ، ويتوقف الحل على ما إذا كانت العينة قد اختيرت بحيث تواجه تكلفة إجمالية محددة C أو تعطي تبايناً محدداً V لـ $\bar{y}_{st}$ وإذا كانت التكلفة مثبتة، فنعوض القيم المثلى لـ n_h في دالة التكلفة (5.17) ثم نحلّ لإيجاد n وهذا يعطي

$$n = \frac{(C - c_0) \sum (N_h S_h / \sqrt{c_h})}{\sum (N_h S_h \sqrt{c_h})} \tag{5.24}$$

وإذا كان V مثبتاً، فنعوض القيمة المثلى n_h في العلاقة الخاصة بـ $V(\bar{y}_{st})$ لنجد:

$$n = \frac{\left(\sum W_h S_h \sqrt{c_h}\right) \sum W_h S_h / \sqrt{c_h}}{V + (1/N) \sum W_h S_h^2} \tag{5.25}$$

حيث $W_h = N_h / N$.

وتبرز حالة خاصة مهمة إذا كان $c_h = c$ أي إذا بقيت التكلفة لكل وحدة نفسها في جميع الطبقات. فالتكلفة تصبح $C = c_0 + cn$ ،وتصبح المحاصّة المثلى في حالة تكلفة مثبتة هي المحاصّة المثلى من أجل حجم عيّنة ثابت. وتكون النتيجة في هذه الحالة الخاصة كما يلي.

نظرية (٥ ـ ٧)

في معاينة عشوائية طبقية يكون $V(\bar{y}_{st})$ أصغر ما يمكن من أجل حجم كلي مثبت للعيّنة n إذا كان

$$n_h = n \frac{W_h S_h}{\sum W_h S_h} = n \frac{N_h S_h}{\sum N_h S_h} \tag{5.26}$$

وتدعى هذه المحاصّة أحياناً محاصّة نيمان، على اسم مبتكرها Neyman (1934) الذي منح برهانـه شهـرة لهذه النتيجة. وقد اكتُشف فيما بعد وجود برهان أقدم بوساطة Tschuprow (1923) .

ونحصـل على علاقـة التبـاين الأصغـري مع n مثبت بتعـويض قيمة n_h من (5.26) في العلاقة العامة المتعلقة بـ $V(\bar{y}_{st})$ وتكون النتيجة

$$V_{min}(\bar{y}_{st}) = \frac{\left(\sum W_h S_h\right)^2}{n} - \frac{\sum W_h S_h^2}{N} \tag{5.27}$$

ويمثل الحد الثاني في الطرف الأيمن عامل الـ ت م م .

(٥ - ٦) الدقة النسبية لمعاينة عشوائية طبقية ومعاينة عشوائية بسيطة

إذا استُخدمت طريقة التقسيم إلى طبقات بمهارة فإنها ستُنتج على الدوام، تقريباً، تبايناً لتقدير المتوسط أو تقدير المجموع أصغر من التباين الذي تعطيه العيّنة العشوائية البسيطة المقابلة. وعلى أي حال، فإنه ليس صحيحاً أن أي عينة عشوائية طبقية تعطي تبايناً أصغر مما تعطيه العيّنة العشوائية البسيطة. وإذا كانت قيم n_h بعيدة عن كونها مثلى، فقد يكون للمعاينة الطبقية تباين أعلى. وفي الحقيقة، وكما سنبيّن، فإنه في حالة حجم كلي مثبت للعيّنة، يمكن حتى للتقسيم إلى طبقات مع محاصّة مثلى، أن يعطي تبايناً أعلى، علماً أن هذه النتيجة تبدو نوعاً من الفضول الأكاديمي أكثر مما هي شيء يُحتمل حدوثه في الممارسة العملية.

ونقوم في هذه الفقرة بمقارنة بين العينة العشوائية البسيطة والمعاينة العشوائية الطبقية بحصص متناسبة أومثلى. وتساعد هذه المقارنة في إظهار كيفية إنجاز الكسب العائد لاستخدام طريقة التقسيم إلى طبقات.

ونرمز لتباينات تقديرات المتوسط بـِ V_{ran} ، V_{prop} ، و V_{opt} على الترتيب.

نظرية (٥ - ٨)

إذا تجاهلنا الحدود $1/N_h$ بالمقارنة مع الواحد، يكون

$$V_{opt} \leq V_{prop} \leq V_{ran} \qquad (5.28)$$

حيث تتم المحاصّة المثلى من أجل n مثبّتة، أي بحيث يكون $n_h \alpha N_h S_h$

برهان

$$V_{ran} = (1-f)\frac{S^2}{n} \qquad (5.29)$$

$$V_{prop} = \frac{(1-f)}{n}\sum W_h S_h^2 = \frac{\sum W_h S_h^2}{n} - \frac{\sum W_h S_h^2}{N} \qquad (5.30)$$

[من المعادلة (5.8) الفقرة (٥-٣)]

$$V_{opt} = \frac{\left(\sum W_h S_h\right)^2}{n} - \frac{\sum W_h S_h^2}{N} \tag{5.31}$$

[من المعادلة (5.27) الفقرة (٥-٥)]

ومن المطابقة الجبرية المعتادة لتحليل تباين المجتمع المقسَّم إلى طبقات، نحصل على

$$\begin{aligned}(N-1)S^2 &= \sum_h \sum_i (y_{hi} - \bar{Y})^2 \\ &= \sum_h \sum_i (y_{hi} - \bar{Y}_h)^2 + \sum_h N_h(\bar{Y}_h - \bar{Y})^2 \\ &= \sum_h (N_h - 1)S_h^2 + \sum_h N_h(\bar{Y}_h - \bar{Y})^2 \end{aligned} \tag{5.32}$$

وبما أن الحدود التي تحوي $1/N_h$ مهملة وبالتالي أيضاً الحدود التي تحوي $1/N$ فالعلاقة (5.32) تعطي

$$S^2 = \sum W_h S_h^2 + \sum W_h(\bar{Y}_h - \bar{Y})^2 \tag{5.33}$$

ومنه

$$V_{ran} = (1-f)\frac{S^2}{n} = \frac{(1-f)}{n}\sum W_h S_h^2 + \frac{(1-f)}{n}\sum W_h(\bar{Y}_h - \bar{Y})^2 \tag{5.34}$$

$$= V_{prop} + \frac{(1-f)}{n}\sum W_h(\bar{Y}_h - \bar{Y})^2 \tag{5.35}$$

ومن تعريف V_{opt} يجب أن يكون $V_{prop} > V_{opt}$. والفرق بينهما هو بموجب العلاقتين (5.30) و (5.31) :

$$\begin{aligned}V_{prop} - V_{opt} &= \frac{1}{n}\left[\sum W_h S_h^2 - \left(\sum W_h S_h\right)^2\right] \\ &= \frac{1}{n}\left[\sum W_h (S_h - \bar{S})^2\right]\end{aligned} \tag{5.36}$$

حيث $\bar{S} = \sum W_h S_h$ هو متوسط مرجّح للمقادير S_h .

ومن (5.35) و (5.36) مع إهمال الحدود في $1/N_h$ ، نجد

$$V_{ran} = V_{opt} + \frac{1}{n}\sum W_h(S_h - \bar{S})^2 + \frac{(1-f)}{n}\sum W_h(\bar{Y}_h - \bar{Y})^2 \qquad (5.37)$$

وخلاصة القول، تبيّن (5.37) أن التباين يتناقص وفق مركبتين عندما ننتقل من المعـاينـة العشوائية البسيطة إلى المحاصّة المثلى. وتأتي المركبة الأولى (الحد الموجود في أقصى اليمين) من حذف الفروق بين متوسطات الطبقات، كما تأتي المركبة الثانية (الحد الأوسط من الطرف الأيمن) من حذف التأثيرات الناتجة عن الفروق بين الانحرافات المعيارية للطبقات. وتمثل المركبة الثانية الفرق في التباين بين المحاصّة المثلى والمحاصّة التناسبية.

وإذا لم يكن ممكناً إهمال الحدود في $1/N_h$ فإن تعويض S^2 من (5.32) يقود إلى النتيجة

$$V_{ran} = V_{prop} + \frac{(1-f)}{n(N-1)}\left[\sum N_h(\bar{Y}_h - \bar{Y})^2 - \frac{1}{N}\sum (N-N_h)S_h^2\right] \qquad (5.38)$$

بدلاً من (5.35) .

ومنـه فإن المعـاينة العشوائية التناسبية تعطي تبايناً أعلى من المعاينة العشوائية البسيطة إذا كان

$$\sum N_h(\bar{Y}_h - \bar{Y})^2 < \frac{1}{N}\sum (N-N_h)S_h^2 \qquad (5.39)$$

ورياضياً، يمكن أن يحدث هذا، فلنفرض أن جميع المقادير S_h^2 تساوي S_w^2 بحيث تكون المحاصّة التناسبية مثلى بالمعنى النيماني للكلمة، فعندئذ تصبح (5.39)

$$\sum N_h(\bar{Y}_h - \bar{Y})^2 < (L-1)S_w^2$$

أو

$$\frac{\sum N_h(\bar{Y}_h - \bar{Y})^2}{L-1} < S_w^2 \qquad (5.40)$$

وأولئك الذين يألفون تحليل التباين سيتعرفون على حقيقة أن هذه العلاقة تتضمن كون متوسط مربعات ما بين الطبقات أصغر من متوسط مربعات ما ضمن الطبقات، أي أن النسبة F أقل من الواحد.

(٥ ـ ٧) متى يُنتج التقسيم إلى طبقات مكاسب كبيرة في الدقة؟

المتغير النموذجي الذي يمكن استخدامه للتقسيم إلى طبقات هو قيمة المتغير y نفسه - أي المقدار الذي نقيسه في عملية المسح. وإذا استطعنا التقسيم إلى طبقات بوساطة قيم y فسوف لا يوجد تداخل بين الطبقات، وسيكون تباين ما ضمن الطبقات أصغر بكثير من التباين الإجمالي، خاصة إذا كان عدد الطبقات كبيراً. وقد أوضحت هذه الحالة بمثال في الفقرة (٥ ـ ٣). صفحة ... وقد تألّف المجتمع من حجوم (أعداد السكان) 64 مدينة عام 1930، مقسّمة إلى طبقات وفق حجمها. ومع وجود طبقتين فقط، فقد خفضت المعاينة الطبقية التناسبية $s.e(\hat{Y})$ من 2365 إلى 1372 والتقسيم إلى طبقات مع $n_1=n_2=12$ هو التقسيم الأمثل وفق المحاصّة النيهانية، يؤدي إلى تخفيض إضافي إلى 1044.

وعمليًا، لا يمكننا بالطبع التقسيم إلى طبقات بوساطة قيم y، إلا أن بعض التطبيقات المهمة تقترب من هذه الحالة، وبالتالي تعطي مكاسب كبيرة في الدقة، من خلال تحقيقها للشروط الثلاثة التالية:

١ ـ المجتمع مؤلف من مؤسسات تتغير تغيّراً واسعاً من حيث حجمها.

٢ ـ المتغيّرات الرئيسة التي سنقيسها مرتبطة ارتباطاً وثيقاً بحجوم المؤسسات.

٣ ـ يتوافر لنا قياس جيد نعتمد عليه لإقامة الطبقات.

وكأمثلة نذكر الأعمال من نوع محدد مثل البقاليات (في مسوح تتعلق بحجم العمل أو عدد المستخدمين)، المدارس (في مسوح تتصل بأعداد الطلاب)، المستشفيات (في دراسات تتعلق بعدد المرضى)، وعائدات ضريبة الدخل (لمفردات ترتبط ارتباطاً عالياً بدخل غير معفى من الضرائب) وفي الولايات المتحدة تتغير المزارع أيضاً تغيراً كبيراً من حيث حجمها مقاساً بمساحتها بالفدادين أو بدخلها الإجمالي، إلا أن بعض مفردات المزرعة المعروفة مثل إنتاج محاصيل معينة أو أنواع من الحيوانات، تُظهر في الغالب ارتباطاً معتدلاً فقط مع حجم المزرعة، وهكذا لا تكون المكاسب الناتجة عن التقسيم إلى طبقات وفقاً لحجم المزرعة مكاسب ضخمة.

وإذا بقي حجم المؤسسة مستقراً عبر الزمن، على الأقل لفترات قصيرة، فيكون أفضل قياس عملي لها هو عادة حجمها وفقاً لتعداد عام مأخوذ في مناسبة حديثة العهد. ويوضح المثال في الفقرة (٥-٣) حالة توافرت فيها معلومات سابقة جيدة. فالجدول (٥-٢) يبيّن الـ S_h والقيم المثلى الناتجة لِـ n_h ($n_h \propto N_h S_h$) وذلك عندما نقوم بالمحاصّة من بيانات 1920 و 1930 على الترتيب.

وتعطي بيانات 1920 قيمة لِـ n_1 تساوي 11.56 في مقابل قيمة مثلى صحيحة مساوية 12.21 لبيانات 1930 . وعند التدوير إلى أعداد صحيحة تعطي كلا المجموعتين من البيانات المحاصة نفسها ـ أي حجم عيّنة قدره 12 في كل طبقة.

جدول (٥-٢) حساب المحاصة المثلى

الطبقة	N_h	بيانات 1920			بيانات 1930		
		S_h	$N_h S_h$	n_h	S_h	$N_h S_h$	n_h
1	16	163.30	2612.80	11.56	232.04	3712.64	12.21
2	48	58.55	2810.40	12.44	74.71	3586.08	11.79
المجاميع	64		5423.20	24.00		7298.72	24.00

ونلاحظ أن كسر المعاينة الأمثل هو 75% في الطبقة الأولى إلا أنه 25% فقط في الطبقة الثانية. وغالباً ما نجد أنه بسبب التباين المرتفع للطبقة المؤلفة من المؤسسات الأكبر، فالعلاقة تستدعي معاينة نسبتها 100% في هذه الطبقة. وفي الحقيقة، يمكن أن تستدعي المحاصّة معاينة بنسبة أكثر من 100% (انظر الفقرة ٥ ـ ٨). لاحظ أيضاً أن الـ S_h في 1920 أصغر منها في 1930 . وتعطي بيانات 1920 انطباعاً متفائلاً للغاية عن الدقة التي سنحصل عليها من مسح 1930 . وكما ذكرنا في الفقرة (٤-٧) فإنه ينبغي أن نأخذ في اعتبارنا دائماً إمكانية تغيّر في مستويات الـ S_h عند استخدام بيانات سابقة، مع أن الهامش المتروك لمثل هذا التغيير قد يكون بالضرورة نوعاً من الحزر أو التخمين.

والتقسيم الجغرافي إلى طبقات، حيث تكون الطبقات مساحات متراصة مثل نواح أو أحياء في مدينة، هو أمر شائع ـ وغالباً ما يكون ذلك توخياً للسهولة من الناحية

الإدارية أو بسبب أن العديد من العوامل تؤثر في اتجاه جعل المحاصيل النامية، أو البشر القاطنين، في المنطقة نفسها تُظهر تشابهاً في خواصها الرئيسة. وعلى سبيل المثال، يبيّن الجدول (٥ ـ ٣) بيانات نشرها Jessen (1920) و Jessen مع Houseman (1944) حول فعالية التقسيم الجغرافي إلى طبقات وذلك في حالة عدد من المفردات الاقتصادية النموذجية لمزرعة.

وقد عُرضت أربعة حجوم للطبقة ـ البلدة، المقاطعة، منطقة يسودها نوع من النشاط الزراعي، والولاية. ولإعطاء فكرة ما عن الحجوم النسبية للطبقة نذكر أنه توجد 1600 بلدة، 100 مقاطعة، و 5 مناطق زراعية في أيوا.

وفي الجدول نعتبر دقة طريقة في التقسيم متناسبة عكسياً مع قيمة $V(\bar{y}_{st})$ المعطاة بهذه الطريقة. وهكذا تكون الدقة النسبية للطريقة 1 إلى الطريقة 2 هي النسبة $V_2(\bar{y}_{st})/V_1(\bar{y}_{st})$ معبراً عنها على شكل نسبة مئوية. والبيانات المعطاة هنا هي متوسطات مأخوذة فوق أعداد المفردات المدوّنة في العمود الثاني. وفي كل حالة أخذنا المقاطعة كشيء قياسي. وكما نرى فإن المكاسب في الدقة معتدلة. وفي أيوا نجد أن استخدام 1600 طبقة (بلدة) بالمقارنة مع عدم التقسيم إلى طبقات (ولاية) يزيد الدقة بحوالي 30% ؛ أي أنها تخفضّ التباين بنسبة حوالي 25% .

جدول (٥ ـ ٣) الدقة النسبية لأنواع مختلفة من التقسيم الجغرافي إلى طبقات (بالنسبة المئوية)

ولاية	عدد المفردات	طبقة: بلدة	مقاطعة	منطقة يسودها نوع من النشاط الزراعي	ولاية
أيوا (1938)	18	115	100	96	91
أيوا (1939)	19	121	100	97	91
فلوريدا (1942)					
منطقة حمضيات	14	144	100	..	
منطقة خضروات	15	111	100	..	
كاليفورنيا (1942)	17	113	100	97	

وفيما يتعلق بالتقسيم الطبقي المتناسب في مقابل التقسيم الأمثل، هناك حالتان يفوز فيهما التقسيم الأمثل فوزاً ميسّراً. الأولى هي الحالة التي ناقشناها لتونّا، وفيها يتألف المجتمع من مؤسسات كبيرة وصغيرة، مقسمة وفق قياس ما للحجم. والتباينات S_h^2 عادة أكبر بكثير في المؤسسات الكبرى منها في المؤسسات الصغرى، مما يجعل التقسيم المتناسب غير فعّال. ونعثر على الحالة الثانية في مسوح تكون معها المعاينة من بعض الطبقات أعلى بكثير من المعاينة من طبقات أخرى وتأثير العوامل $\sqrt{C_h}$ قد يجعل أداء المحاصة التناسبية ضعيفاً.

وعند تخطيط محاصّة لا تختلف n_h المقدّرة فيها اختلافاً كبيراً عن المحاصة التناسبية يكون من الجدير أن نُقدِّر التضخم في مقدار $V(\bar{y}_{st})$ أو $V(\hat{y}_{st})$ في حال استخدام المحاصّة التناسبية. والحل الأمثل لمسألة المحاصّة هو إلى حد ما متعدد الإمكانيات أو القيم (يمتد على مدى فترة معينة)، [انظر الفقرة (٥ا-٢)]. وقد نجد الزيادة في التباين صغيرة بصورة مدهشة. وفضلاً عن ذلك، فإن تفوّق الحل الأمثل كما حسبناه من القيم المقدّرة لـ S_h مبالغ فيه دائماً وذلك بسبب الأخطاء المرتكبة في تقدير S_h. وقد تستحق بساطة المحاصّة التناسبية وميزة الترجيح الذاتي فيها زيادة 10 إلى 20% في التباين.

(٥ ـ ٨) المحاصّة التي تحتاج إلى معاينة تزيد على ١٠٠٪

وكما ذكرنا في الفقرة (٥ ـ ٧)، قد تُنتج العلاقة الخاصة بالحل الأمثل قيمة لـ n_h في بعض الطبقات أكبر من N_h الموافقة. لنعتبر المثال المتعلق بحجوم المدن في الفقرة (٥ ـ ٣). فقد استدعت العينة من 24 مدينة، موزعة بين طبقتين، أن نأخذ 12 مدينة من بين 16 في الطبقة الأولى، و 12 من بين 48 في الطبقة الثانية. ولو كان حجم العيّنة 48 فستتطلب المحاصّة أن نأخذ 24 مدينة من أصل 16 في الطبقة الأولى. وأفضل ما يمكن القيام به هو أن نأخذ كل المدن في الطبقة، تاركين 32 من المدن للطبقة الثانية بدلاً من الـ 24 التي تفترضها العلاقة. وتبرز هذه المشكلة فقط عندما يكون كسر المعاينة الإجمالي كبيراً، وتكون إحدى الطبقات أكثر تغيراً بكثير من الطبقات الأخرى. وقد حدث ذلك في مناسبات متعددة.

وإذا أعطت المحاصّة الأصلية $n_1>N_1$ في حالة وجود أكثر من طبقتين، فالمحاصة المثلى المحسّنة تكون :

$$\tilde{n}_1 = N_1; \qquad \tilde{n}_h = (n - N_1)\frac{W_h S_h}{\sum_2^L W_h S_h}, \qquad (h \geq 2) \tag{5.41}$$

شريطة أن يكون $\tilde{n}_h \leqslant N_h$ لكل $h\geqslant 2$ وإذا حدث أن كان $\tilde{n}_2>N_2$ فنغيّر المحاصّة إلى

$$\tilde{n}_1 = N_1; \qquad \tilde{n}_2 = N_2; \qquad \tilde{n}_h = (n - N_1 - N_2)\frac{W_h S_h}{\sum_3^L W_h S_h}, \tag{5.41'}$$

شريطة أن يكـون $\tilde{n}_h \leqslant N_h$ لكـل $h\geqslant 3$. ونستمر بهذه الطريقة حتى يكون $\tilde{n}_h \leqslant N_h \tilde{n}$ لكل ويمكن البرهان على أن المحاصّة الناتجة مثلى من أجل n معطى، كما هو متوقع .

ويجب اتخاذ الحذر بحيث نستخدم العلاقة الصحيحة لِـ $V(\bar{y}_{St})$ وتكون العلاقة العامة (5.6) في الفقرة (٥ ـ ٣) صحيحة إذا عوضنا فيها $\tilde{n}_h$ التي تعطيها المحاصّة المثلى المحسنّة . أما العلاقة (5.27) الخاصة بِـ $V_{min}(\bar{y}_{st})$, وهي :

$$V_{min}(\bar{y}_{st}) = \frac{(\sum W_h S_h)^2}{n} - \frac{\sum W_h S_h^2}{N} \tag{5.27}$$

فلا تعود صحيحة . وإذا رمزنا بِـ $\sum'$ للمجموع فوق الطبقات التي يكون فيها $\tilde{n}_h \leqslant N_h$ فتكون العلاقة البديلة الصحيحة هي

$$V_{min}(\bar{y}_{st}) = \frac{(\sum' W_h S_h)^2}{n'} - \frac{\sum' W_h S_h^2}{N} \tag{5.42}$$

حيث n' هو المجموع الكلي المحسنّ للعيّنة في هذه الطبقات .

(٥ ـ ٩) تقدير حجم العيّنة في حالة معلومات البيانات المتصلة

قدمنا في الفقرة (٥ ـ ٥) علاقات لتحديد n تحت محاصّة مثلى في تقديرنا . ونقدم في هذه الفقرة علاقات لأي محاصّة، مع بعض الحالات الخاصة المفيدة . لقد افترضنا هناك أن للتقدير تبايناً محدداً V فلنحدد بدلاً من ذلك هامش الخطأ d (فقرة ٤ ـ ٤)،

$V=(d/t)^2$ حيث t قيمة المتغير الطبيعي الموافقة للاحتمال المسموح به لحادثة تجاوز الخطأ الفعلي للهامش المرغوب.

تقدير متوسط مجتمع $\bar{Y}$:

ليكن s_h تقدير S_h و $w_h=w_h n$ حيث الـ w_h مقادير تمَّ اختيارها. وبدلالة هذه الحدود يكون $V(\bar{y}_{st})$ المنتظر على الشكل (من النظرية ٥ـ٣، فقرة ٥ـ٣):

$$V=\frac{1}{n}\sum\frac{W_h^2 s_h^2}{w_h}-\frac{1}{N}\sum W_h s_h^2 \tag{5.43}$$

حيث $W_h=N_h/N$ وهذا يعطي كعلاقة عامة في n :

$$n=\frac{\sum\frac{W_h^2 s_h^2}{w_h}}{V+\frac{1}{N}\sum W_h s_h^2} \tag{5.44}$$

وإذا تجاهلنا عامل الـ ت م م نجد كتقريب أول

$$n_0=\frac{1}{V}\sum\frac{W_h^2 s_h^2}{w_h} \tag{5.45}$$

وإذا كان n_0/N غير مهمل فيمكن حساب n على الشكل

$$n=\frac{n_0}{1+\frac{1}{NV}\sum W_h s_h^2} \tag{5.46}$$

وفي حالات خاصة يمكن أن تأخذ هذه العلاقات أشكالاً حسابية أكثر سهولة. ونعطي هنا القليل منها.

محاصّة مثلى افتراضية (n مثبت): $w_h \propto W_h s_h$.

$$n=\frac{(\sum W_h s_h)^2}{V+\frac{1}{N}\sum W_h s_h^2} \tag{5.47}$$

محاصّة تناسبية :

$$w_h = W_h = N_h/N.$$

$$n_o = \frac{\sum W_h s_h^2}{V}, \qquad n = \frac{n_0}{1+\frac{n_0}{N}} \tag{5.48}$$

تقدير مجموع المجتمع

إذا كان V هو القيمة المرغوبة لِـ $V(\hat{Y}_{st})$ فإن العلاقات الرئيسة تصبح كالتالي :

الشكل العام :

$$n = \frac{\sum \frac{N_h^2 s_h^2}{w_h}}{V + \sum N_h s_h^2} \tag{5.49}$$

مثلى افتراضية :

$$n = \frac{(\sum N_h s_h)^2}{V + \sum N_h s_h^2} \tag{5.50}$$

تناسبية :

$$n_0 = \frac{N}{V} \sum N_h s_h^2, \qquad n = \frac{n_0}{1+\frac{n_0}{N}} \tag{5.51}$$

مثال

هذا المثـال مأخـوذ من بحث لِـ Cornel (1947) يدرس فيه عينـة من كليات جامعـات الـولايات المتحدة مسحوبة عام 1946 من قبل مكتب التربية في الولايات المتحدة، وذلك لتقدير عدد المسجلين للعام الدراسي 1947-1946 . والتوضيح هنا يتعلق بمجتمع يتضمن 169 من الكليات ودور المعلمين. وقد رُتّبت هذه في سبع طبقات، سنهمل طبقة صغيرة منها. وقد شُكلت الطبقات الخمس الأولى وفقاً لحجم المعهد. وتضمنت السادسة كليات البنات فقط وحُسبت s_h(تقديرات الـ S_h) من نتائج العام الدراسي 1944-1943 . وقد استُخدم تقسيم «أمثل» للطبقات مبني على هذه القيم لِـs_h.

وكـان الهـدف هو معـامـل اختـلاف مقـداره 5% في تقـدير العـدد الإجمـالي للمسجلين. وفي عام 1943 كان عدد المسجلين الإجمالي لهذه المجموعة من الكليات

56472 . وهكذا يكون الخطأ المعياري المرغوب

$$(0.05)\ (56472) = 2824$$

بحيث يصبح التباين المرغوب

$$V = (2824)^2 = 7,974,976$$

وقد يكون هناك اعتراض مفاده أن التسجيل في 1946 سيكون أكبر مما هو في 1943 وأن هامشاً يجب أن يترك لهذه الزيادة . وفي الواقع ، فإن الحسابات تفترض فقط أن معامل الاختلاف للكلية الواحدة يبقى كما هو في عامي 1943 أو 1946 ، وهو فرض قد لا يكون مجانباً للمنطق .

ويبيّن الجدول (٥-٤) قيم n_h ، s_h و $N_h S_h$ التي كانت معروفة قبل تحديد n .
والعلاقة المناسبة لتحديد n هي العلاقة (5.50) التي تنطبق على محاصّة «مثلى» تهدف إلى تقدير مجموع . ومن غير المحتمل أن يكون عامل الـ (ت م م) مهملاً في مجتمع كهذا لا يحوي إلا 196 وحدة . وعلى أي حال، وبغية التوضيح ، سنحسب أول تقريب متجاهلين الـ ت م م . وهو

$$n_0 = \frac{(\sum N_h s_h)^2}{V} = \frac{(26,841)^2}{7,974,976} = 90.34$$

جدول (٥-٤) بيانات من أجل تقدير حجم عيّنة

طبقة	N_h	s_h	$N_h s_h$	n_h
1	13	325	4,225	9
2	18	190	3,420	7
3	26	189	4,914	10
4	42	82	3,444	7
5	73	86	6,278	13
6	24	190	4,560	10
مجاميع	196		26,841	56

ومن الــواضـح أننـا في حاجـة إلى التعـديل. فمن أجـل القيمـة الصحيحـة لِـ n في (5.50) نجد

$$n = \frac{n_0}{1+\frac{1}{V}\sum N_h s_h^2} = \frac{90.34}{1+\frac{4,640,387}{7,974,976}} = 57.1$$

وقد اختير حجم عيّنة مساو لِـ 56* وتظهر قيم الـ n_h للطبقات كل بمفردها في العمود الأيمن من الجدول (٥ـ٤).

(٥ - ١٠) المعاينة الطبقية في حالة النسب

إذا رغبنا في تقدير نسبة الوحدات في المجتمع التي تقع في صف معيّن C. فإننا نصـل إلى تقسيم نموذجي للطبقات عندما نستطيع أن نضع في الطبقة الأولى جميع الوحدات التي تقع في C وفي الثانية كل وحدة لا تقع في C، وإذ نفشل في بلوغ ذلك، نحاول أن نشكل الطبقات بحيث تختلف نسبة الوحدات من الصف C بالقدر الممكن من طبقة إلى أخرى.

لتكن

$$P_h = \frac{A_h}{N_h}, \qquad p_h = \frac{a_h}{n_h}$$

نسب الوحدات من الصف C في الطبقة h وفي العيّنة المأخوذة من هذه الطبقة، على الترتيب. فتقدير النسبة في كامل المجتمع الموافق لمعاينة عشوائية طبقية هو:

$$p_{st} = \sum \frac{N_h p_h}{N} \tag{5.52}$$

نظرية (٥ـ٩)

في المعاينة العشوائية الطبقية يكون تباين p_{st} معطى بالعلاقة:

$$V(p_{st}) = \frac{1}{N^2}\sum \frac{N_h^2(N_h - n_h)}{N_h - 1}\frac{P_h Q_h}{n_h} \tag{5.53}$$

---

* تختلف النتائج الحسابية قليلاً عن تلك التي أعطاها Cornell (1947)

برهان

هذه حالة خاصة من النظرية العامة المتعلقة بتباين المتوسط المقدّر. ومن النظرية (٥ ـ ٣) لدينا

$$V(\bar{y}_{st}) = \frac{1}{N^2}\sum N_h(N_h - n_h)\frac{S_h^2}{n_h} \qquad (5.54)$$

ليكن y_{hi} متغيراً قيمته 1 عندما تقع الوحدة في C وصفرًا فيما عدا ذلك، ففيما يتعلق بهذا المتغيّر بيّنا في المعادلة (3.4) من الفقرة (٣ ـ ٢) أن

$$S_h^2 = \frac{N_h}{N_h - 1}P_hQ_h \qquad (5.55)$$

وهذا يؤدي إلى النتيجة المطلوبة.

ملاحظة: في جميع التطبيقات، وحتى لو لم يكن عامل الـ ت م م مهملاً، فستكون الحدود التي تحوي $1/N_h$ مهملة عمليًا، وبالتالي يمكن استخدام العلاقة الأبسط إلى حد ما

$$V(p_{st}) = \frac{1}{N^2}\sum N_h(N_h - n_h)\frac{P_hQ_h}{n_h} = \sum \frac{W_h^2P_hQ_h}{n_h}(1 - f_h) \qquad (5.56)$$

نتيجة (١)

عندما نستطيع تجاهل عامل الـ ت م م نكتب

$$V(p_{st}) = \sum W_h^2\frac{P_hQ_h}{n_h} \qquad (5.57)$$

نتيجة (٢)

في حالة المحاصّة التناسبية

$$V(p_{st}) = \frac{N-n}{N}\,\frac{1}{nN}\sum\frac{N_h^2P_hQ_h}{N_h - 1} \qquad (5.58)$$

$$\doteq \frac{1-f}{n}\sum W_hP_hQ_h \qquad (5.59)$$

ولتقـدير التبـاين، من العيّنـة، يمكن تعـويض $p_hq_h/(n-1)$ بدلاً من الكمية المجهولة P_hQ_h/n_h في أي من العلاقات المذكورة أعلاه.

وينتـج أفضـل اختيار للمقادير n_h التي تجعل $V(p_{st})$ أصغر ما يمكن من النظرية العامة في الفقرة (٥-٥).

والتباين الأصغري في حالة قيمة مثبتة للحجم الكلي للعيّنة هو

$$n_h \propto N_h\sqrt{N_h/(N_h-1)}\sqrt{P_hQ_h} \doteq N_h\sqrt{P_hQ_h}$$

وهكذا فإن

$$n_h \doteq n\frac{N_h\sqrt{P_hQ_h}}{\sum N_h\sqrt{P_hQ_h}} \qquad (5.60)$$

التباين الأصغري الموافق لتكلفة مثبتة، حيث التكلفة $= c_0+\sum c_hn_h$

$$n_h \doteq n\frac{N_h\sqrt{P_hQ_h/c_h}}{\sum N_h\sqrt{P_hQ_h/c_h}} \qquad (5.61)$$

ويتم إيجاد قيمة n كما في الفقرة (٥-٥).

(٥-١١) المكاسب في الدقة في حالة معاينة طبقية للنسب

إذا كانت التكاليف على أساس الوحدة الواحدة هي نفسها في جميع الطبقات، فهناك قاعدتا عمل مفيدتان (أ) يكون الكسب في الدقة من المعاينة الطبقية العشوائية فوق المعاينة العشوائية البسيطة صغيراً أو متواضعاً ما لم تتغير النسب P_h تغيراً كبيراً من

جدول (٥-٥) الدقة النسبية للمعاينة العشوائية الطبقية والبسيطة

الدقة النسبية (%)	طبقية $nV(p_{st})/(1-f) = \frac{1}{3}\sum P_hQ_h$	بسيطة $nV(p)/(1-f) = PQ$	P_h
103	2433	2500	0.4, 0.5, 0.6
112	2233	2500	0.3, 0.5, 0.7
132	1900	2500	0.2, 0.5, 0.8
174	1433	2500	0.1, 0.5, 0.9

طبقة إلى طبقة، و(ب) إذا وقعت جميع النسب P_h بين 0.1 و 0.9 فإن كسب المحاصّة المثلى في حالة n مثبت فوق المحاصّة التناسبية يكون كسباً بسيطاً.

ولتوضيح النتيجة الأولى، يقارن الجدول (٥ـ٥) معاينة عشوائية طبقية (محاصّة تناسبية) مع معاينة عشوائية بسيطة من أجل ثلاث طبقات بالحجم نفسه $(W_h=\frac{1}{3})$ ونجد أربع حالات: الأولى فيها P_h تساوي 0.4 ,0.5 ,0.6 في الطبقات الثلاث، أما الأخيرة (وهي الأكثر تطرفاً) ففيها P_h تساوي 0.1 ,0.5 ,0.9 . ويبيّن العمودان التاليان جداء تباينات النسبة المقدّرة بـ $n/(1-f)$ ويعطي العمود الأخير الدقة النسبية للمعاينة العشوائية الطبقية إلى المعاينة العشوائية البسيطة. والكسب في الدقة هو كسب كبير في الحالتين الأخيرتين فقط.

ولمقارنة المحاصتين التناسبية والمثلى في حالة n مثبت، سنجد أنه مع إهمال العامل $(1-f)$:

$$V_{opt}=\frac{(\sum W_h\sqrt{P_hQ_h})^2}{n}, \qquad V_{prop}=\frac{\sum W_hP_hQ_h}{n} \tag{5.62}$$

وهكذا تصبح الدقة النسبية للمحاصّة التناسبية فوق المحاصّة المثلى هي:

$$\frac{V_{opt}}{V_{prop}}=\frac{(\sum W_h\sqrt{P_hQ_h})^2}{\sum W_hP_hQ_h} \tag{5.63}$$

وإذا وقعت جميع المقادير P_h بين القيمتين P_0 و $(1-P_0)$ فإننا نهتم بالقيمة الصغرى التي تأخذها الدقة النسبية. وللتبسيط، نأخذ طبقتين بحجمين متساويين $W_1=W_2$. فنبلغ الدقة النسبية الصغرى عندما يكون $P_1=\frac{1}{2}$ و $P_2=P_0$. وعندئذ تصبح الدقة النسبية

$$\frac{V_{opt}}{V_{prop}}=\frac{(0.5+\sqrt{P_0Q_0})^2}{2(0.25+P_0Q_0)} \tag{5.64}$$

وبعض قيم هذه الدالّة معطاة في الجدول (٥ـ٦). وحتى في حالة كون $P_0=0.1$ أو كونها مرتفعة إلى الحد 0.9 فإن الدقة النسبية هي 94 بالمائة. وفي معظم الحالات تكون ميزتا

البساطة والترجيح الذاتي للمحاصّة التناسبية أكثر من أن تعوّض هذه الخسارة البسيطة في الدقة .

وينبغي التنويه بمحدودية هذا المثال . فهو لا يأخذ في الاعتبار الفروق بين تكاليف المعاية في الطبقات المختلفة . وفي بعض المسوح تكون النسب P_h صغيرة جدًّا ولكنها تتراوح ، مثلاً ، بين 0.001 و 0.05 في الطبقات المختلفة . وقد تتحقق هنا مكاسب مرموقة أكثر من التقسيم الأمثل إلى طبقات .

جدول (٥-٦) جدول الدقة النسبية للمحاصّتين المثلى والتناسبية

P_0	0.4 or 0.6	0.3 or 0.7	0.2 or 0.8	0.1 or 0.9	0.05 or 0.95
RP(%)	100.0	99.8	98.8	94.1	86.6

(٥-١٢) تقدير حجم العيّنة في حالة النسب

ويمكن استنتاج القوانين المتعلقة بتحديد حجم عيّنة من القوانين الأكثر شمولاً في الفقرة (٥-٩) . ليكن V التباين المرغوب لتقدير النسبة P المتعلقة بالمجتمع ككل ، فالقوانين المتعلقة بالنوعين الرئيسين للمحاصّة هي كما يلي :

تناسبية :

$$n_0 = \frac{\sum W_h p_h q_h}{V}, \qquad n = \frac{n_0}{1 + \frac{n_0}{N}} \tag{5.65}$$

المثلى المفترضة :

$$n_0 = \frac{\left(\sum W_n \sqrt{p_h q_h}\right)^2}{V}, \qquad n = \frac{n_0}{1 + \frac{1}{NV}\sum W_h p_h q_h} \tag{5.66}$$

حيث n_0 هو التقريب الأول، الذي يتجاهل عامل الـ ت م م . و n هي القيمة المصححة آخذين في الاعتبار عامل الـ ت م م . وعند اشتقاق هذه القوانين اعتبرنا العامل $N_h/(N_h-1)$ مساوياً للواحد .

تنطبق كل نتائج هذه الفقرة على تقدير نسبة ما. وإذا كان من المفضّل استخدام النسب المئوية، فالعلاقات نفسها ممكنة التطبيق إذا عبّرنا عن V, Q_h, P_h إلخ. على شكل نسب مئوية. ولتقدير العدد الكلي من وحدات المجتمع التي تقع في الصف C أي تقدير P فإننا نضرب كل التباينات بـ N^2.

تمارين

(٥-١) في مجتمع فيه $N=6$ و $L=2$ كانت قيم y_{hi} هي 0, 1 في الطبقة 1 و 4، 6 و 11 في الطبقة 2. ونريد أخذ عيّنة حجمها 2.

(أ) بيّن أن المحاصّة المثلى النيمانية، عند التدوير إلى عدد صحيح، هي $n_h=1$ في الطبقة 1 و $n_h=3$ في الطبقة 2.

(ب) احسب $\overline{y}_{St}$ لكل عيّنة ممكنة يمكن سحبها تحت المحاصّة المثلى وتحت المحاصة التناسبية.

تحقّق من أن التقديرين غير منحازين. وبالتالي أوجد $V_{opt}(\overline{y}_{st})$ و $V_{prop}(\overline{y}_{st})$ مباشرة.

(ج) تحقّق من أن $V_{opt}(\overline{y}_{st})$ يتفق مع العلاقة المعطاة في (5.6) ومن أن $V_{prop}(\overline{y}_{st})$ يتفق مع العلاقة المعطاة في (5.8) صفحة ١٣٦.

(د) استخدم العلاقة (5.27) صفحة ـ لحساب $V_{opt}(\overline{y}_{st})$ الذي يتضمن خطأ طفيفاً لأنه لا يفسح المجال لحقيقة أن الـ n_h مقربة إلى أقرب عدد صحيح. هل تتفق هذه النتيجة جيداً مع القيمة المصححة.

(٥-٢) نأخذ عيّنة من مجموع الأسر في مدينة لتقدير الكمية الوسطية لممتلكات الأسرة الواحدة التي يمكن ردّها مباشرة إلى معادل نقدي. وقد قُسمت الأسر إلى طبقتين عالية الإيجار ومنخفضة الإيجار. ويُعتقد أن بيتاً من طبقة الإيجار العالي يحوي حوالي 9 أمثال ما يحويه بيت من طبقة الإيجار المنخفض من مثل هذه الممتلكات، كما يُتوقع أن يكون S_h متناسباً مع الجذر التربيعي لمتوسط الطبقة. ويوجد 4000 من الأسر في طبقة الإيجار المرتفع و 20000 في طبقة الإيجار المنخفض.

(أ) كيف توزع عيّنة من 1000 أسرة بين الطبقتين؟

(ب) كيف ينبغي توزيع العيّنة إذا كان الهدف هو تقدير الفرق بين ممتلكات الأسرة الواحدة في الطبقتين؟

(٥-٣) تبيّن المعلومات الإحصائية التالية تقسيم جميع المزارع في منطقة إلى طبقات وفقاً لحجم المزرعة، ولمتوسط عدد الفدادين من الذُرة في المزرعة ضمن كل طبقة.

حجم المزرعة بالفدان	عدد المزارع N_h	متوسط فدادين الذرة $\bar{Y}_h$	الانحراف المعياري S_h
0–40	394	5.4	8.3
41–80	461	16.3	13.3
81–120	391	24.3	15.1
121–160	334	34.5	19.8
161–200	169	42.1	24.5
201–240	113	50.1	26.0
241–	148	63.8	35.2
المجموع أو المتوسط	2010	26.3	

في عيّنة حجمها 100 مزرعة احسب حجوم العيّنات من كل طبقة تحت (ا) المحاصّة التناسبية. (ب) المحاصّة المثلى. قارن دقة كل من هاتين الطريقتين بدقة المعاينة العشوائية البسيطة.

(٥-٤) برهن النتيجة المعروضة في العلاقة (5.38) فقرة (٥-٦):

$$V_{ran} = V_{prop} + \frac{(1-f)}{n(N-1)}\left[\sum N_h(\bar{Y}_h - \bar{Y})^2 - \frac{1}{N}\sum (N-N_h)S_h^2\right]$$

(٥-٥) لدى معاين طبقتان حجماهما النسبيان W_1 ، W_2 ويعتقد أنه يمكن اعتبار S_1 ، S_2 متساويين إلا أنه يُظن أن C_2 قد تكون بين $2C_1$ و $4C_1$ ويُفضل استخدام محاصة تناسبية، إلا أنه لا يرغب في التعرض لزيادة كبيرة في التباين بالمقارنة مع المحاصّة المثلى. في حالة تكلفة معطاة $C=c_1n_1+c_2n_2$ ومتجاهلاً الـ ت م م. بيّن أن

$$\frac{V_{prop}(\bar{y}_{st})}{V_{opt}(\bar{y}_{st})} = \frac{W_1c_1 + W_2c_2}{(W_1\sqrt{c_1} + W_2\sqrt{c_2})^2}$$

إذا كان $W_1=W_2$ فاحسب الـزيادة النسبية في التبـاين عند استخدام المحاصّة التناسبية، وذلك عندما $c_2/c_1=2,4$.

(٥-٦) يقـترح معـاين أخذ عينة عشوائية طبقية، ويتوقع أن تكون التكاليف الميدانية وفق الصيغة $\sum c_h n_h$ وكانت تقديراته المسبقة عن الحجمين المناسبين للطبقتين كما يلي:

Stratum	W_h	S_h	C_h
1	0.4	10	$4
2	0.6	20	$9

(ا) أوجد قيم n_1/n و n_2/n التي ستجعل التكلفة الميدانية الإجمالية أصغر ما يمكن وذلك في حالة قيمة معطاة لـ $V(\bar{y}_{st})$.

(ب) أوجـد حجم العيّنـة المـطلوب تحت هذه المحـاصّـة التنـاسبية، كي تجعل $V(\bar{y}_{st})=1$. تجاهل الـ ت م م.

(ج) كم ستكون التكلفة الإجمالية للعمل الميداني؟

(٥-٧) بعد أخذ العيّنة في التمرين (٥-٦) يجد المعاين أن التكلفة الفعلية للعمل الميداني هي 2$ لكل وحدة من الطبقة 2

(ا) ما هي زيادة التكلفة الميدانية عما توقّعه؟

(ب) لو أنـه عرف التكـاليف الميدانية الصحيحـة سلفاً، هل كان سيتمكن من بلوغ $V(\bar{y}_{St})=1$ من أجـل التكلفـة الميدانية المقّـدرة في الأصـل في التمرين (٥-٦)؟ (تلميح: تعـطي متراجحـة كوشي - شوارتز صفحة ١٤٢، حيث $V'=1$ جواب هذا السؤال دون إيجاد المحاصة الجديدة).

(٥ - ٨) في عملية تقسيم إلى طبقتين كانت قيم S_h و W_h كما يلي

Stratum	W_h	S_h
1	0.8	2
2	0.2	4

احسب حجمي العيّنتين n_1 ، n_2 اللتين نحتاجهما من الطبقتين لتحقيق الشروط التالية (وتتطلب كل حالة حسابات منفصلة؛ تجاهل الـ ت م م):

(ا) نريد خطأ معيارياً مساوياً 0.1 لتقدير متوسط المجتمع كما نريد حجم عيّنة كلّيًا $n=n_1+n_2$ أصغر ما يمكن.

(ب) نريد خطأ معيارياً مساوياً 0.1 لتقدير المتوسط في كل طبقة.

(ج) نريد خطأ معيارياً مساوياً 0.1 للفرق بين تقديري متوسطي العيّنتين في الطبقتين، على أن يكون حجم العيّنة الكلي أصغر ما يمكن أيضاً.

(٥-٩) مع وجود طبقتين، يرغب المعاين، متوخياً السهولة في الأعمال الإدارية، في أخذ $n_1=n_2$ بدلاً من استخدام القيم التي تعطيها المحاصّة النيمانية. إذا كان $V_{opt}(\bar{y}_{st}), V(\bar{y}_{st})$ ، يرمزان للتباينين الناتجين عن $n_1=n_2$ والمحاصّة النيمانية، على الترتيب، بيّن أن الزيادة الكسرية في التباين

$$\frac{V(\bar{y}_{st})-V_{opt}(\bar{y}_{st})}{V_{opt}(\bar{y}_{st})}=\left(\frac{r-1}{r+1}\right)^2$$

حيث $r=n_1/n_2$ كما تعطيها المحاصّة النيمانية. ومن أجل الطبقتين في التمرين (٥ ـ ٨)، حالة (ا)، كم ستكون الزيادة الكسرية في التباين عند استخدام $n_1=n_2$ بدلاً من المحاصّة المثلى؟

(٥-١٠) إذا كانت دالة التكلفة من الشكل $C=c_0+\sum t_h\sqrt{n_h}$، حيث c_0 و t_h أعداد معروفة، بين أنه كي يكون $V(\bar{y}_{st})$ أصغر ما يمكن في حالة تكلفة إجمالية مثبتة فإن n_h يجب أن يتناسب مع

$$\left(\frac{W_h{}^2S_h{}^2}{t_h}\right)^{2/3}$$

أوجد الـ n_h لحجم عيّنة مقداره 1000 وذلك تحت الشروط التالية:

طبقة	W_h	S_h	t_h
1	0.4	4	1
2	0.3	5	2
3	0.2	6	4

(٥-١١) إذا كان $V_{prop}(\bar{y}_{st})$ تباين تقدير المتوسط من عيّنة عشوائية بسيطة حجمها n مع محاصّة تناسبية و $V(\bar{y})$ هو تباين عيّنة عشوائية بسيطة حجمها n ، فبيّن أن النسبة

$$\frac{V_{prop}(\bar{y}_{st})}{V(\bar{y})}$$

لا تعتمد على حجم العيّنة إلا أن النسبة

$$\frac{V_{min}(\bar{y}_{st})}{V_{prop}(\bar{y}_{st})}$$

تتناقص عندما يزداد n (ويتضمن هذا أن المحاصّة المثلى في حالة n مثبت تصبح أكثر فعالية بالنسبة إلى المحاصّة التناسبية كلما ازداد n) [استخدم العلاقتين (5.8) و (5.27)]

(٥-١٢) قارن القيمتين اللتين نحصل عليهما لـ $V(p_{St})$ تحت المحاصّة التناسبية والمحاصّة المثلى في حالة حجم عيّنة مثبت في كل من المجتمعين التاليين. حجوم الطبقات متساوية. ويمكن إهمال الـ ت م م .

المجتمع ١		المجتمع ٢	
طبقة	P_h	طبقة	P_h
1	0.1	1	0.01
2	0.5	2	0.05
3	0.9	3	0.10

ما هي النتيجة العامة التي يوضحها هذان المجتمعان؟

(٥-١٣) بيّن أنه في تقدير النسب تصبح النتائج المقابلة للنظرية (٥-٨) كما يلي:

$$V_{ran} = V_{prop} + \frac{(1-f)}{n} \sum W_h (P_h - P)^2$$

$$V_{prop} = V_{opt} + \frac{\sum W_h (\sqrt{P_h Q_h} - \overline{\sqrt{P_h Q_h}})^2}{n}$$

حيث

$$\overline{\sqrt{P_h Q_h}} = \sum W_h \sqrt{P_h Q_h}.$$

(٥-١٤) نعلم أن 62% من المستخدمين في شركة هم ذكور مَهَرة أو غير مَهرة. و 31% من الكَتبة إناث، و 7% يعملون في الرقابة. ومن عينة من 400 مستخدم ترغب الشركة في تقدير نسبة المستخدمين الذين يستفيدون من تسهيلات استجمامية معينة. وتقول التخمينات التقريبية الأولى إن التسهيلات تُستخدم من قبل 40% إلى 50% من الذكور، 20% إلى 30% من الإناث، و 5% إلى 10% من المراقبين.

(أ) كيف تقترح تقسيم العينة بين الزّمر الثلاث؟

(ب) إذا كانت النسب الصحيحة للمستجمّين 48%, 21%, 4% على الترتيب، فماذا ستكون الأخطاء المعيارية للنسبة المقدّرة؟

(ج) ماذا سيكون الخطأ المعياري لِـ p من عيّنة عشوائية بسيطة حجمها $n=400$ ؟

(٥ - ١٥) تحت المحاصّة النيمانية تصبح العلاقة (5.27) الخاصة بالتباين الأصغري لِـ $\bar{y}_{St}$ كما يلي:

$$V_{min}(\bar{y}_{st}) = \frac{(\sum W_h S_h)^2}{n} - \frac{\sum W_h S_h^2}{N}$$

ويعلّق طالب كما يلي: «بما أن $\sum W_h S_h^2 > (\sum W_h S_h)^2$ ما لم تكن المقادير S_h كلها متساوية، فلابد أن تكون العلاقة خاطئة باعتبار أنه عندما يقترب n من N ستعطي العلاقة قيمة سالبة لِـ $V(\bar{y}_{st})$.» هل الطالب مخطىء أم العلاقة؟

(٥ ـ ١٦) من العلاقة نجد أن كسر المعاينة في الطبقة h هو في المحاصّة اليَنهانية $f_h = n_h/N_h = nS_h/N \sum W_j S_j$. والحالات التي تستدعي فيها هذه العلاقة نسبة معاينة تفوق الـ 100% من طبقة ما $(f_h > 1)$ هي إذن في الغالب تلك التي يكون كسر المعاينة الإجمالي n/N فيها كبيرًا إلى حد ما، ويكون التباين في إحدى الطبقات كبيرًا بصورة غير عادية. وفيما يلي مثال مجتمع صغير، حيث $n=40, N=100$

طبقة	N_h	S_h	الأمثل n_h
1	60	2	15
2	30	4	15
3	10	15	10
	100		40

(ا) تحقق من أن القيم المثلى لِـ n_h هي كما نرى في العمود الأيمن.

(ب) احسب $V(\bar{y}_{st})$ مستخدمًا العلاقة (5.6) ثم العلاقة (5.24) وبيّن أن كلتيهما تعطي $V(\bar{y}_{st})=0.12$.

إضافات في أوجه المعاينة الطبقيّة

(١٥-١) تأثيرات الانحرافات عن المحاصّة المثلى

يناقش هذا الفصل عددًا من الموضوعات الخاصة في مجال الاستخدام العملي للمعاينة الطبقية. وتعالج الفقرات (١٥ - ١) إلى (١٥ - ٨)، (١٥ - ١٠) و(١٥ - ١٥) مسائل يمكن أن تبرز في تخطيط العيّنة: وتعالج الفقرات الباقية طرقًا لتحليل النتائج. ونناقش في هذه الفقرة الخسارة في الدقة الناتجة عن الفشل في تحقيق محاصة مثلى للعيّنة.

لنفرض أننا نعتزم استخدام محاصة مثلى في حالة n معطاة. فينبغي أن يكون حجم العيّنة n_h في الطبقة h

$$n_h' = \frac{n(W_h S_h)}{\sum W_h S_h} \qquad (5A.1)$$

ومن المعادلة (5.27) صفحة ١٤٤، نجد أن التباين الأصغري الناتج هو

$$V_{min}(\bar{y}_{st}) = \frac{1}{n}(\sum W_h S_h)^2 - \frac{1}{N}\sum W_h S_h^2 \qquad (5A.2)$$

ومادام الانحراف المعياري S_h غير معروف عمليًا فيمكننا فقط تقريب هذه المحاصّة. وإذا كان $\hat{n}_h$ هو حجم العيّنة المستخدمة في الطبقة h فإن التباين الذي نبلغه فعلاً من المعادلة (5.6) صفحة ١٣٥، هو

$$V(\bar{y}_{st}) = \sum \frac{W_h^2 S_h^2}{\hat{n}_h} - \frac{1}{N}\sum W_h S_h^2 \qquad (5A.3)$$

وزيادة التباين الناتجة عن المحاصّة التقريبية هي،

$$V(\bar{y}_{st}) - V_{min}(\bar{y}_{st}) = \sum \frac{W_h^2 S_j^2}{\hat{n}_h} - \frac{1}{n}(\sum W_h S_h)^2 \quad (5A.4)$$

لنعوض من *(5A.1)* عن $W_h S_h$ ، بدلالة n_h في الحد الأول من الطرف الأيمن، فهذا يعطي النتيجة المرجوّة

$$V(\bar{y}_{st}) - V_{min}(\bar{y}_{st}) = \frac{(\sum W_h S_h)^2}{n^2}\left(\sum \frac{n_h'^2}{\hat{n}_h} - n\right)$$

$$= \frac{(\sum W_h S_h)^2}{n^2} \sum \frac{(\hat{n}_h - n_h')^2}{\hat{n}_h} \quad (5A.5)$$

وبالعودة إلى المعادلة (5A.2) مهملين الـ ت م م (الحد الأخير من الطرف الأيمن) نجد

$$\frac{V_{min}(\bar{y}_{st})}{n} = \frac{(\sum W_h S_h)^2}{n^2} \quad (5A.6)$$

وبالتالي فإن الزيادة النسبية في التباين الناتجة من الانحرافات عن المحاصة المثلى هي

$$\frac{V(\bar{y}_{st}) - V_{min}(\bar{y}_{st})}{V_{min}(\bar{y}_{st})} = \frac{1}{n} \sum_{h=1}^{L} \frac{(\hat{n}_h - n_h')^2}{\hat{n}_h} \quad (5A.7)$$

حيث n'_h هو الحجم الأمثـل و$\hat{n}_h$ الحجم الفعـلي للعيّنـة في الطبقة h. وإذا لم يكن الـ ت م م مهملاً تصبح إشارة المساواة في (5A.7) متراجحة $\geqslant$.

وليكن $g_h = |\hat{n}_h - n_h'|/\hat{n}_h$ الفرق المطلق بين حجمي العيّنة في الطبقة h معبراً عنه على شكل كسر من الحجم الفعلي $\hat{n}_h$. فتصبح (5A.7) عندئذ

$$\frac{V - V_{min}}{V_{min}} = \sum_{h=1}^{L} \frac{\hat{n}_h}{n} g_h^2 \quad (5A.8)$$

وهــو متــوسط مرجّــح للمقــادير g_h^2 . ولــذلــك يشكـل g^2 حدًّا أعـلى متحفـظًا لِـ $(V\text{-}V_{min})/V_{min}$ ، حيث g هو الفـرق النسبي الأعـلى في جميع الـطبقـات . وهكذا لايمكن أن تتجاوز الزيادة النسبية في التباين $(0.2)^2$ أو 4% إذا كان $g=0.2$ أو 20% وإذا كان $g=30\%$ فالـزيادة النسبية في التبـاين هي على الأكثر 9% . وبهذا المعنى يمكن وصف الأمثلية بأنها مفهوم مَرن يتّسع للعديد من الحلول.

وفضلاً عن ذلك فإن (5A.8) تقترح أن الحد الأعلى g^2 سيبالغ غالبًا مبالغة غير قليلة في تقدير زيادة التباين. ويعطي الجدول (١٥ ـ ١) مثالاً يتضمن ثلاث طبقات حيث $n=340$ وتقضي المحاصة المثلى أن تكون حجوم العيّنات 200 ,100 و 40 بينما الحجوم الفعلية المستخدمة هي 150 ,120 و 70 .

جدول (١٥ ـ ١) تأثيرات الانحراف عن المحاصة المثلى

طبقة	n_h' مثلى	$\hat{n}_h$ فعلية	$\frac{\lvert\hat{n}_h - n_h'\rvert}{\hat{n}_h}$	$\frac{(\hat{n}_h - n_h')^2}{\hat{n}_h}$
1	200	150	0.33	16.7
2	100	120	0.17	3.3
3	40	70	0.43	12.9
المجموع	340	340	—	32.9

وبما أن قيمة g هي 0.43 (طبقة ٣)، فالقاعدة التقريبية تعطي 18% كزيادة نسبية في التباين. ومن العمود الأيمن نرى أن الزيادة الفعلية هي $32.9/340 = 9.7\%$.

وقد درس Evans (1951) المسألة نفسها بدلالة تأثيرات الأخطاء المرتكبة في تقديرات S_h وطوّر قاعدة تقريبية تبيّن ما إذا كان يمكن لمحاصّة مثلى تقديرية أن تكون أدق من محاصّة تناسبية. ويفترض أن معامل الاختلاف لتقدير S_h يبقى نفسه في جميع الطبقات. ويكون هذا الفرض مناسبًا عندما يجري تقدير الـ S_h من عيّنات ابتدائية لها الحجم نفسه في كل طبقة. ويبيّن كيفية حساب الحجم الذي نحتاجه للعيّنة الابتدائية كي تكون المحاصّة «المثلى» أفضل، في المتوسط، من المحاصّة التناسبية. وقبل ذلك كان Sukhatme (1935) قد بيّن أن عيّنة ابتدائية صغيرة تجعل عادة احتمال تفوق المحاصّة «المثلى» على المحاصّة التناسبية احتمالاً مرتفعًا. [انظر Sukhatme و Sukhatme (1970) ، صفحة 88].

(١٥ ـ ٢) تأثيرات أخطاء في حجوم الطبقات

في نوع مرغوب من المعاينة الطبقية، قد لاتكون المجاميع الكلية للطبقات N_h معروفة بالضبط، باعتبارها مأخوذة من بيانات تعداد إحصائي عتيق. وبدلاً من النسب الصحيحة W_h في كل طبقة تتوافر لنا تقديرات w_h . وتقدير العيّنة لِـ $\bar{Y}$ هو عندئذ $\sum w_h \bar{y}_h$.

وبعبارات عامة تكون تبعات استخدام ترجيحات خاطئة كما يلي :

١ ـ تقدير العيّنة منحاز. وبسبب الانحياز نقيس دقة التقدير مستخدمين متوسط مربعات الخطأ حول $\bar{Y}$ بدلاً من تباين التقدير حول متوسطه [انظر الفقرة (١ ـ ٩)].

٢ ـ يبقى الانحياز ثابتًا عندما يزداد حجم العيّنة. وبالتالي، يمكن دائمًا الوصول إلى حجم عيّنة يكون التقدير معه أقل دقة من المعاينة العشوائية البسيطة، ونخسر كل المكاسب في الدقة العائدة إلى التقسيم إلى طبقات.

٣ ـ التقدير المعتاد $s(\bar{y}_{st})$ يقلل من تقدير الخطأ الحقيقي في $\bar{y}_{st}$ باعتبار أنه لا يتضمن مساهمة الانحياز في الخطأ.

ولتبرير هذه العبارات نلاحظ أنه عند تكرار العيّنة تكون القيمة المتوسطة للتقدير هي $\sum w_h \bar{Y}_h$. وبالتالي تكون قيمة الانحياز

$$\sum (w_h - W_h) \bar{Y}_h$$

وهي مستقلة عن حجم العيّنة. ولإيجاد متوسط مربعات خطأ (MSE) التقدير يمكن التحقق بسهولة من أن عبارة التباين معطاة بالعلاقة المعتادة، حيث نضع w_h بدلاً من W_h وبالتالي نجد

$$\text{MSE}(\bar{y}_{st}) = \sum \frac{{w_h}^2 {S_h}^2}{n_h}(1 - f_h) + [\sum (w_h - W_h) \bar{Y}_h]^2 \qquad (5A.9)$$

وهذه العبارة من إنتاج Stephan (1941) وأخيراً نقول إن العلاقة المعتادة لـ $s^2(\bar{y}_{st})$ هي بوضوح تقدير غير منحاز للحد الأول من (5A.9) إلا أنها تأخذ في الاعتبار الحد الثاني.

مثال

يوضح هذا المثال الخسارة في الدقة الناتجة عن ترجيحات غير صحيحة وذلك عندما يكون التقسيم إلى طبقات

(ا) فعـالاً بصورة طفيفة، (ب) فعالاً بصورة مرتفعة. لنعتبر مجتمعًا كبيرًا فيه $S^2=1$ ويمكن تقسيمـه إلى طبقتين حيث $W_1 = 0.9, \; W_2 = 0.1$. وسنفترض $S_1 = S_2 = S_h$. فعندئذ وبإهمال حدود في $1/N_h$ نجد

$$S^2 \doteq \sum W_h S_h^2 + \sum W_h (\bar{Y}_h - \bar{Y})^2 \tag{5A.10}$$
$$= S_h^2 + W_1 W_2 (\bar{Y}_1 - \bar{Y}_2)^2$$

أي أن

$$1 = S_h^2 + 0.09(\bar{Y}_1 - \bar{Y}_2)^2$$

وفي (ا) خذ $\bar{Y}_1 - \bar{Y}_2 = 1.$ فعنـدئـذ $S_h^2 = 0.91,$ والمحاصّة التناسبية بترجيحات صحيحة تخفض التباين بنسبة 9% بالمقارنة مع المعاينة العشوائية البسيطة.

وفي (ب) خذ. $\bar{Y}_1 - \bar{Y}_2 = 3,$ مما يعطي $S_h^2 = 0.19,$ أي تخفيض في التباين يزيد على 80% . ومع طبقتين وترجيحات غير صحيحة، يمكن كتابة الانحياز على الشكل

$$(w_1 - W_1)(\bar{Y}_1 - \bar{Y}_2)$$

باعتبـار أن $(w_1 - W_1) = -(w_2 - W_2)$.لنفرض أن الترجيحات المقدّرة هي $w_1=0.92$ و$w_2=0.08$ فيصل الانحياز إلى 0.02 = 1 (0.02) في (ا) وإلى 0.06 في (ب). وبالتالي نجد المقادير التالية لِـ MSE في حالة حجم للعيّنة مساوٍ لِـ n .

معاينة عشوائية بسيطة: $V(\bar{y}) = \dfrac{1}{n}$

معاينة عشوائية طبقية:

$$(a): \quad \text{MSE}(\bar{y}_{st}) = \frac{0.91}{n} + 0.0004$$

$$(b): \quad \text{MSE}(\bar{y}_{st}) = \frac{0.19}{n} + 0.0036$$

وكما يبيّن الجدول (٥ا-٢)، تبدأ المعاينة العشوائية البسيطة في أن تكون الرابحة بالمقارنة مع (ا) عند $n=300$. وعلى أي حال فهناك القليل من الاختيار بين الطريقتين حتى حدود $n=1000$.

وفي (ب)، حيث الأمر أكثر خطورة، تبقى المعاينة الطبقية متفوقة حتى $n=200$ مع أن معظم ما كسبناه يكون قد فُقد عند هذا الحد لحجم العيّنة. ووراء $n=300$ تصبح المعاينة الطبقية متخلفة بصورة ملحوظة عن المعاينة العشوائية البسيطة. ويكون التقدير الـدقيق للترجيحـات w_h مهمًّا بصورة خاصة عندما يكون التقسيم إلى طبقات مرتفع الفعالية أو عندما يكون حجم العينة كبيراً.

جدول (١٥-٢) قيم مقارنة لمتوسط مربعات خطأ $(\bar{y})$

n	عشوائية بسيطة	عشوائية طبقية (a)	عشوائية طبقية (b)
50	0.0200	0.0186	0.0074
100	0.0100	0.0095	0.0055
200	0.0050	0.0049	0.0045
300	0.0033	0.0034	0.0042
400	0.0025	0.0027	0.0041
1000	0.0010	0.0013	0.0038

وفي بعض المسوح الإحصائية يمكن أخذ حجم كبير n' لعيّنة ابتدائية كي نقدّر الـ w_h. ولهذه الطريقة، وتُعرف بالمعاينة المضاعفة، أو المعاينة ثنائية الطور، تطبيقات عديدة ونناقشها في الفصل ١٢- وسنبرهن أنه في حالة المعاينة المضاعفة يكون متوسط مربعات الخطأ لِـ $\bar{y}_{St}$ مساويًا تقريبًا:

$$\frac{\sum W_h S_h^2}{n} + \frac{\sum W_h(\bar{Y}_h - \bar{Y})^2}{n'} \qquad (5A.11)$$

وبمقارنة متوسط مربعات الخطأ مع S^2/n كما تعطيه العلاقة (5A.10) نرى أننا نحتفظ بمعظم الكسب من التقسيم إلى طبقـات شريطة أن يكـون n' أكـبر بكثـير من n. ولـوضـع الفكرة في صورة أعم نقـول إن مجموعة من الترجيحات المقدّرة تحفظ معظم الكسب الكبير للمعـاينة الطبقية إذا كان تقديرها أدق بكثير مما لو أُخذت من عيّنة عشوائية بسيطة حجمها n.

(١٥ ـ ٣) مسألة المحاصّة في حالة أكثر من مفردة واحدة

بما أن المحاصة الأفضل لمفردة معينة سوف لا تكون، بصورة عامة، الأفضل أيضًا لمفردة أخرى، فلا بد من الوصول إلى حل وسط في مسح إحصائي يحوي عدة مفردات. والخطوة الأولى هي أن نختزل المفردات التي سنأخذها في اعتبارنا عند المحاصّة إلى عدد صغير نسبيًا يعتقد أنها المفردات الأكثر أهمية. وإذا توافر لنا بيان إحصائي سابق جيد، فيمكن عندئذ حساب المحاصّة المثلى لكل مفردة على حدة، ورؤية مدى عدم التوافق بينها. وفي مسح إحصائي من نوع خاص، يمكن أن يكون الارتباط بين المفردات مرتفعًا، وقد تختلف عندئذ المحاصّات اختلافًا طفيفًا نسبيًا.

مثال

يوضح البيان الإحصائي لِـ Jessen (1942) مسح مزارع من هذا النوع. فقد قسمت ولاية أيوا إلى خمس مناطق جغرافية، نرمز لكل منها بنشاطها الزراعي الرئيس. لنفرض أن هذه المناطق ستُستخدم كطبقات في مسح إحصائي يتعلق بمزارع إنتاج الحليب ومشتقاته. والمفردات الثلاث الأكثر أهمية هي عدد البقرات المحلوبة في اليوم الواحد، عدد جالونات الحليب في اليوم، والدخل السنوي من الأموال التي يتم استلامها لقاء منتجات الحليب ومشتقاته. ومن مسح جرى عام 1938 نرى في الجدول (١٥ـ٣) التقديرات للانحرافات المعيارية ضمن الطبقات. ويقدم الجدول (١٥ـ٤) المحاصّة النيهانيّة المثلى على أساس هذه القيم لِـ s_h لكل مفردة على حدة وذلك في عيّنة من 1000 مزرعة.

جدول (١٥ـ٣) الانحرافات المعيارية ضمن الطبقات

الطبقة	$W_h = \frac{N_h}{N}$	s_h البقرات المحلوبة	s_h جالونات من الحليب	s_h إيصالات بأثمان منتجات ألبان (بالدولار)
Northeast dairy	0.197	4.6	11.7	332
Cash grain	0.191	3.4	9.8	357
Western livestock	0.219	3.3	7.0	246
Southern pasture	0.184	2.8	6.5	173
Eastern livestock	0.208	3.7	9.8	279

جدول (١٥-٤) حجوم العيّنة ضمن الطبقات $n=1000$.

الطبقة	المحاصّة				
	تناسبية	مثلى من أجل			متوسط m_h
		أبقار	جالونات	إيصالات	
Northeast dairy	197	254	258	236	250
Cash grain	191	182	209	246	212
Western livestock	219	203	171	194	189
Southern pasture	184	145	134	115	131
Eastern livestock	208	216	228	209	218

ولا تختلف المحـاصّـات المثـلى بعضها عن بعض اختلافات حادة. وباستثناء واحد، تنحرف الأرقام للمفردات الثلاث في اتجاه واحد عن المحاصّة التناسبية. وهكذا نجد المحاصّة التناسبية تقترح في الطبقة الأولى 197 من المزارع بينما تقود المحاصّات لكل من المفـردات إلى أرقام بين 236 و 258 ويقدّم متوسط الحجوم المثلى للعيّنات الخاصة بالمفردات الثلاث، وهو المتوسط الذي يظهر في العمود الأيمن. محاصّة وسطية مُرضية.

جدول (١٥ ـ ٥) التباينات المتوقعة للمتوسط التقديري

نوع المحاصّة	أبقار	جالونات	مبالغ
مثلى	0.0127	0.0800	76.9
وسطية	0.0127	0.0800	76.9
تناسبية	0.0131	0.0837	80.9

ويبيّن الجدول (١٥-٥) التباينات المتوقعة للمتوسط التقديري $\bar{y}_{st}$ كما ينتج من المحاصّات الثلاث المثلى والوسطيّة والتناسبية، كل بمفردها، والعلاقات التي تُعطي هذه التباينات هي:

$$v_{opt} = \frac{(\sum W_h s_h)^2}{n}, \qquad v_{comp} = \sum \frac{(W_h S_h)^2}{m_h}, \qquad v_{prop} = \frac{\sum W_h s_h^2}{n}$$

وتقدم المحاصّة الوسطية نتائج لها، تقريبًا، الدقة نفسها كما لو كان ممكنًا أن نستخدم محاصّات مثلى منفصلة لكل مفردة. وما تجدر ملاحظته هو أن دقة المحاصّة التناسبية لا تقل عن دقة المحاصّة الوسطية والمحاصات المثلى المنفردة إلا بمقدار بسيط، وفضلاً عن ذلك، فإن الجدول (١٥ ـ ٥) يبالغ في تقدير دقة المثلى والوسطية باعتبارهما تَمتا باستخدام تباينات تقديرية بدلاً من التباينات الصحيحة. وتشكل هذه النتيجة توضيحًا آخر لمرونة مفهوم الأمثلية المذكور في الفقرة (١٥ـ١).

(١٥ ـ ٤) طرق أخرى للمحاصّة في حالة أكثر من مفردة واحدة

اقترح Chatterjee (1967) محاصّة وسطية بديلة وهي أن نختار الحجوم n_h التي تجعل معدل الزيادات النسبية في التباين وفق (5A.7) أصغر ما يمكن، على أن يؤخذ المعدّل فوق المتغيرات المختلفة. وإذا رمزنا للمتغير بِـ j فإن هذا يؤدي إلى اختيار

$$n_h = n\sqrt{\sum_j n_{jh}'^2}\Big/\sum_h\sqrt{\sum_j n_{jh}'^2} \qquad (5A.12)$$

حيث n'_{jh} هو حجم العيّنة الأمثل في الطبقة h الموافق للمتغير j. وفي البيان المذكور في الجـدول (١٥ـ٤)، حيث تختلف المحـاصّات المثلى بصورة طفيفة فقط، تختلف قيم Chatterjee لِـ n_h عن المعدّل m_h في الجدول (١٥ـ٤) بوحدة واحدة على الأكثر في أي طبقة.

وفي بعض المسوح تختلف المحاصات المثلى الموافقة للمتغيرات المختلفة بشكل كبير، وبحيث لا يوجد حل وسطي واضح. ونحتاج إلى مبدأ ما نحدد بموجبه المحاصة التي سنستخدمها. ونقدّم هنا طريقتين مفيدتين اقترحهما Yates (1960).

وتنطبق الأولى على مسوح لها أهداف خاصة، يمكننا معها قياس الخسارة المترتبة على حجم معطى لخطأ التقدير بدلالة المال أو المنفعة، كما ناقشنا في الفقرة (٤ ـ ١٠). ومع k من المتغيرات ودوال خسارة تربيعية يمكن التعبير، بصورة مقبولة، عن الخسارة

الكلية المتوقعة كدالة خطية في تباينات المتوسطات التقديرية لمجتمع. وفي حالة المتوسطات لدينا

$$L=\sum_{j}^{k} a_j V(\bar{y}_{jst})=\sum_{j}^{k} a_j \sum_{h}^{L} {W_h}^2 S_{jh}^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right) \qquad (5A.13)$$

حيث S^2_{jh} هو تباين المتغير في الطبقة h . وتغيير ترتيب إشارتَي المجموع يعطي

$$L=\sum_{h}\frac{{W_h}^2}{n_h}\left(\sum_{j} a_j S_{jh}^2\right)-\frac{1}{N}\sum W_h\left(\sum_{j} a_j S_{jh}^2\right) \qquad (5A.14)$$

وفي حالة دالة خطية في تكاليف المعاينة نجد

$$C=c_0+\sum c_h n_h \qquad (5A.15)$$

وبجعل جداء $(C-c_0)$ بالحد الأول من L (الحد الذي يعتمد على n_h) أصغر ما يمكن نجد، مستخدمين متراجحة كوشي - شوارتز:

$$n_h \propto \frac{W_h}{\sqrt{c_h}}\sqrt{\sum_{j} a_j S_{jh}^2} \qquad (5A.16)$$

ونعثر على ثابت التناسب بتحقيق القيود المفروضة على L أو C وعلى سبيل المثال، لنفرض أن قيمة L محددة وأنه يمكن تجاهل حد الـ ت م م. فنجد

$$n_h=\frac{n(W_h A_h/\sqrt{c_h})}{\sum (W_h A_h/\sqrt{c_h})} \qquad (5A.17)$$

حيث $A_h=\sqrt{\sum_{j} a_j S_{jh}^2}$. وحجم العيّنة الكلي المطلوب هو وفقًا لـ (5A.14)

$$n=\frac{1}{L}\left(\sum_{h}\frac{W_h A_h}{\sqrt{c_h}}\right)\left(\sum_{h} W_h A_h\sqrt{c_h}\right) \qquad (5A.18)$$

وفي الطريقة الثانية نحدد التباينات المرغوبة V_j لكل متغير. وفي حالة متوسطات المجتمع يتضمن هذا كون

$$\sum_{h=1}^{L}\frac{{W_h}^2 S_{jh}^2}{n_h}-\sum_{h=1}^{L}\frac{W_h S_{jh}^2}{N}\le V_j \qquad (j=1,2,\ldots,k) \qquad (5A.19)$$

وتُستخدم إشارة المتراجحة لأن المحاصة الأكثر اقتصادًا يمكن أن تزودنا بتباينات أصغر من التباين المرغوب V_j وذلك في بعض المفردات.

وبهذه الطريقة نجعل التكلفة C [معادلة (5A.15)] أصغر ما يمكن خاضعة للتساهلات V_j وللشروط $0 \le n_h \le N_h$. والمسألة هي إحدى مسائل البرمجة غير الخطية. وقد وُضعت خوارزميات لحلّها من قبل Hartly وHocking (1963)، Chatterjee (1966)؛ Zukhovitsky وAvdeyeva (1966)؛ Haddleston وآخرون (1970). وكان Dalenius (1957)، قبل ذلك، قد أعطى حلًّا بيانيًّا بارعًا، بينما طوّر Yates (1960) و Kokan (1963) طرق تقريبات متتالية موضحة في الطبعة الثانية من هذا الكتاب.

والخطوة المفيدة الأولى هي بالطبع القيام بمحاصّة مثلى لكل من المتغيرات بمفرده وإيجاد تكلفة تحقيق حد التساهل لكل منها. لنأخذ مثلًا المتغير y_1 ذا التكلفة الأعلى C_1 ولندرس ما إذا كانت القيم المثلى n_h الموافقة لِـ y_1 محققة لجميع شروط التساهل الـ (k-1) الأخرى. فإذا كان الأمر كذلك نستخدم هذه المحاصّة وتكون المسألة قد حُلّت لأن أي محاصّة أخرى سوف لا تحقق التساهل V_1 الخاص بِـ y_1 بتكلفة منخفضة مثل التكلفة C_1

وبحلّ سلسلة من الأمثلة في مسألة من هذا النوع، أشار Boothe و Sedransk (1969) إلى أنه مع فشل البرمجة على الحاسب يمكن الحصول في الغالب على تقريب جيد لحل مسألة Yates الثانية وذلك بحل المسألة الأولى الأسهل. لنحدد أن L في (5A.13) ستأخذ القيمة $V^* = \sum a_j V_j$ حيث المقادير V_j هي التساهلات المرغوبة كل بمفردها وأن القيم a_j متناسبة عكسيًّا مع الـ V_j ففي حالة متغيرين يكون $a_1 = V_2/(V_1 + V_2)$ ، و $a_2 = V_1/(V_1 + V_2)$ ، ويكون:

$$V^* = \frac{2 V_1 V_2}{(V_1 + V_2)} \qquad (5A.20)$$

مثال (أربع طبقات، متغيران)

البيان الإحصائي وتطبيق الطريقة التقريبية مبيّنان في الأعمدة من (1) إلى (6) من الجدول (٥-٦). والمسألة هي إيجاد n الصغرى التي تحقق:

$$V(\bar{y}_{1st}) \leqq 0.04, \qquad V(\bar{y}_{2st}) \leq 0.01$$

جدول (١٥-٦) معلومات إحصائية مصطنعة في حالة أربع طبقات ومتغيرين

العمود	(1)	(2)	(3)	(4)	(5)	(6)	(7)
الطبقة	W_h	S^2_{1h}	S^2_{2h}	$A_h{}^2=\sum_j a_j S^2_{jh}$	$W_h A_h$	n_h	$\tilde{n}_h$
1	0.4	25	1	5.8	0.963	206	194
2	0.3	25	4	8.2	0.859	183	180
3	0.2	25	16	17.8	0.844	180	187
4	0.1	25	64	56.2	0.750	160	171
				المجاميع	3.416	729	732

وباستخدام المحاصّة المثلى لكل متغير بمفرده نجد أنه من السهل التحقق من أننا نحتاج إلى $n=625$ لتحقيق الشرط الأول وإلى $n=676$ لتحقيق الشرط الثاني. إلا أن $n=676$ والمحاصّة الناتجة عنها لا تحقق الشرط الأول إذ تؤدي إلى $V_1=0.0589$ بدلاً من 0.04 وبطريقة التقريبات المتتالية نجد حلاً يحقق كلي الشرطين (الحل مبينّ في الطبعة الثانية) وهو $\tilde{n}=732,$ والمقادير $\tilde{n}_h$ مبينة في العمود (7) من الجدول (١٥ ـ ٦).

ولتطبيق طريقة Booth و Sedransk حيث $V_1=0.04$ ، $V_2 0.01$ نأخذ

$$L = 0.2\,V(\bar{y}_{1st}) + 0.8\,V(\bar{y}_{2st}) = \frac{2(0.04)(0.01)}{(0.05)} = 0.016$$

ومن (5A.18) حيث $c_h=1$ نجد

$$n = \frac{\left(\sum W_h A_h\right)^2}{L} = \frac{(3.416)^2}{0.016} = 729$$

مستخـدمين العمود (5) من الجدول (١٥ ـ ٦). ومن المعادلة (5A.17) نجد أيضًا أن العمـود (5) يقـود إلى قيم n_h المبينـة في العمـود (6) من الجدول (١٥ ـ ٦). وكما يبينّ العمودان (6) و (7) يتفق الحلان n_h و $\tilde{n}_h$ بصورة جيدة.

وكما يلاحظ Booth و Sedransk فإن $L = V^*,$ في جميع المسائل من هذا النوع،

باعتبار أن n يحقق الشرط الوحيد $L=V^*$ ولا حاجة لأن يحقق الشروط الخاصة بكل متغير، بينما تحقق المحاصة $\tilde{n}_h$ الشرط L بالإضافة إلى الشروط الأخرى كل بمفردها.

(١٥ ـ ٥) التقسيم إلى طبقات في اتجاهين مع عيّنات صغيرة

لنفرض وجود قاعدتين للتقسيم إلى طبقات، مثلاً وفقًا لِـ R من الصفوف و C من الأعمدة، مما يؤدي إلى RC من الخلايا. وإذا كان $n \geqslant RC$ فيمكن تمثيل كل خلية في العيّنة. وتبرز مشكلة عندما يكون $n<RC$ ونرغب في أن تمنح العيّنة تمثيلاً متناسبًا لكل من قاعدتي التقسيم إلى طبقات. وفي طريقة بسيطة طوّرها Hartley, Bryant و (1960) Jessen نجد أن حل المشكلة يتطلب فقط أن يتجاوز n أكبر العددين C و R.

ولتوضيح هذه الطريقة، لنفرض أننا نقسم مجتمعًا صغيرًا من 165 مدرسة إلى خمس طبقـات وفقًا لحجم المدينة وإلى أربع طبقات وفقًا لمعدّل الصرف على أساس التلميذ الواحد. ويبيّن الجدول (١٥ـ٧) أعداد المدارس m_{ij} ونسب المدارس $P_{ij} = m_{ij}/165$ في كل من الخلايا العشرين.

والهـدف هو إعطاء كل مدرسة فرصًا متساوية في الاختيار بينما نعطي كل فئة هامشية تمثيلها المتناسب. وفي هذا التوضيح $n=10$. لنحسب الأعداد $n_{i.} = nP_{i.}$ و $n_{.j} = nP_{.j}$، حيث نقرّب هذه الجداءات إلى أقرب عدد صحيح (مع تعديل ثانوي إضافي، في حال الحاجة إليه، بحيث يكون مجموع المقادير n_i مساويًا لِـ n وكذلك مجموع المقادير n_j). وهذه المقادير مبينة في الجدول (١٥ ـ ٧).

والخطوة التالية هي سحب $n=10$ خلايا باحتمال $n_{i.}n_{.j}/n^2$ من أجل الخلية ij. وبذلك نقوم بتشكيل مربع $n \times n$ (جدول ١٥ ـ ٨). ونسحب عمودًا بصورة عشوائية من الصف 1 . ثم نسحب من الصف 2 ، وبصـورة عشوائية، عمودًا من الأعمدة البـاقية، وهلمّ جرّا. وفي النهـاية، كل صف وكل عمود يحوي وحدة. (ونقوم بهذا السحب بأسرع ما يمكن من خلال تباديل عشوائية للأرقام من 1 إلى 10) ونشير إلى نتائج سحبة واحدة بالإشارات × في الجدول (١٥ ـ ٨).

جدول (١٥ ـ ٧) عدد ونسب المدارس في كل خلية

حجم المدينة		المصروف لكل تلميذ A	B	C	D		المجاميع	$n_{i.}$
I	m_{1j} P_{1j}	15 0.091	21 0.127	17 0.103	9 0.055	$m_{1.}$ $P_{1.}$	62 0.376	4
II	m_{2j} P_{2j}	10 0.061	8 0.049	13 0.079	7 0.042	$m_{2.}$ $P_{2.}$	38 0.231	2
III	m_{3j} P_{3j}	6 0.036	9 0.055	5 0.030	8 0.049	$m_{3.}$ $P_{3.}$	28 0.170	2
IV	m_{4j} P_{4j}	4 0.024	3 0.018	6 0.036	6 0.036	$m_{4.}$ $P_{4.}$	19 0.114	1
V	m_{5j} P_{5j}	3 0.018	2 0.012	5 0.030	8 0.049	$m_{5.}$ $P_{5.}$	18 0.109	1
المجاميع	$m_{.j}$ $P_{.j}$ $n_{.j}$	38 0.230 2	43 0.261 3	46 0.278 3	38 0.231 2		165 1.000	

جدول (١٥ ـ ٨) مربع 10 x 10 لسحب عيّنة

الصف		العمود 1	2	3	4	5	6	7	8	9	10
		A			B			C		D	
1	I	×									
2					×						
3			×								
4								×			
5	II						×				
6									×		
7	III			×							
8										×	
9	IV					×					
10	V										×

ونلاحظ أن العمودين 1 و 2 مخصصان للطبقة الهامشية A باعتبار أن $n_{.1}=2$. وبصورة مشابهة، الصفوف 1 إلى 4 مخصصة للطبقة الهامشية I باعتبار أن $n_{1.}=4$ وهكذا. وبذلك نكمل توزيع العيّنة على الخلايا العشرين. وتظهر هذه المحاصّة بشكل أكثر تراصًا في الجدول (١٥-٩). والآن نسحب عشوائيًا مدرستين من المدارس الخمس عشرة في الخلية IA وهلمّ جرًّا. واحتمال سحب مدرسة في الصف i والعمود j متناسب مع $n_{i.}n_{.j}/P_{ij}$ وهكذا تكون الاحتمالات غير متساوية مع أنها ستكون تقريبًا متساوية إذا كان $P_{ij} \doteq n_{i.}n_{.j}/n^2$

وكتقدير غير منحاز للمتوسط على أساس المدرسة الواحدة نجد،

$$\bar{y}_U = \frac{1}{n}\sum \frac{n^2 P_{ij}}{n_{i.}n_{.j}} y_{ij}$$

حيث y_{ij} مجموع العيّنة في الخلية ij وعلى أي حال، إذا كان $P_{ij} \doteq n_{i.}n_{.j}/n^2,$ فقد يكون الأفضل حساب متوسط العينة $\bar{y}$ باعتبار أن انحيازه ينبغي أن يكون في حكم المهمل. ويتوافر لنا من العينة تقدير للتباين في حالة كل من التقديرين المنحاز وغير المنحاز شريطة أن يكون n مساويًا على الأقل لضعف أكبر العددين R و C ، وأن نسحب وحدتين على الأقل من كل صف وكل عمود.

جدول (١٥ - ٩) توزّع العيّنة على الخلايا العشرين

	A	B	C	D	المجموع
I	2	1	1	0	4
II	0	0	2	0	2
III	0	1	0	1	2
IV	0	1	0	0	1
V	0	0	0	1	1
المجموع	2	3	3	2	10

وإذا كان P_{ij} في بعض الخلايا، مختلفًا بصورة ملحوظة عن $n_{i.}n_{.j}/n^2$ فهناك خطوة إضافية تجعل احتمالات اختيار المدارس أكثر تباينًا على وجه التقريب. فبعد حساب الـ n_{ij} و n ندرس المقادير $D_{ij} = nP_{ij} - n_{i.}n_{.j}/n$ بعد تدويرها إلى أقرب عدد

صحيح ، وإذا رأينا ، في أي خلية ، أن D_{ij} عدد صحيح موجب فإننا نخصص بصورة آلية D_{ij} من الوحدات إلى هذه الخلية . ونخفض الآن n والمقادير $n_{i.}$ و $n_{.j}$ بالقدر الذي يستدعيه تخفيض هذه الحصة الثابتة للخلية ij ثم نستمر في المحاصّة على المنوال نفسه .

(١٥ ـ ٦) التحكم في الاختيار

وتسمى طريقـة أخـرى لحل هذه المسـألة في حالة عيّنات صغيرة التحكم في الاختيار وتعود لِـ Goodman و Kish (1950) وقدم Riedel, Hess و Fitzpatrick (1976) توضيحًا بسيطًا يُظهر الفكرة الأساسية في الطريقة عندما طبقوها في معاينة المستشفيات . والاتجاه الرئيسي للتقسيم إلى طبقات هو على أساس حجم المستشفى (طبقتان) . ومن المرغوب فيه أيضًا تمثيل كل من نوعين من مالكي المستشفيات ، إلا أننا سنسحب وحدة (مستشفى) واحدة فقط من كل من الطبقتين الرئيستين . وعند ترقيم الواحدات ضمن الـطبقـات، تشير الفتحة (′) إلى نوع من المالكين بينما يشير غياب الفتحة إلى النوع الآخر. ويبيّن الجانب الأيسر من الجدول (١٥ ـ ١٠) الوحدات في كلتا الطبقتين.

جدول (١٥ـ١٠) ترتيب الوحدات ضمن الطبقات في حالة التحكم في الاختيار

الطبقة بترتيبها الأصلي	الطبقة بترتيبها المحسّن	مستشفى صغير	مستشفى كبير
3′	1	1	1
4′	2	2	2
5′	3′	3′	3′
1	4′	4′	4′
2		5′	

وإذا سحبنا الوحدة 1 أو 2 من الطبقة I ، فنرغب عندئذ في سحب الوحدات 3′، 4′ أو 5′ من الطبقة II . بحيث يكون لكل من اتجاهي التقسيم إلى طبقات حضور في حالة $n=2$. وبصورة مشابهة، 3′ أو 4′ (طبقة I) مرغوبة مع 1 أو 2 (طبقة II). ويجعل التحكم في الاختيار احتمال مثل هذه التراكيب المرغوبة مرتفعًا بالقدر الممكن رياضيًّا، بينما يُبقى احتمالات الاختيار ضمن كل طبقة متساوية، ويؤدي بالتالي إلى

تقديرات غير منحازة من خلال تطبيق العلاقات المعتادة في المعاينة الطبقية. والهدف هو إما زيادة في الدقة في حالة n معطاة أو توفير في التكاليف الميدانية.

وفي حالة معاينة عشوائية طبقية يكون احتمال تركيب مرغوب هو + (.6)(.5) 5. = (.4)(.5) ويمكن زيادة هذا الاحتمال إلى 0.9 من خلال تغييرين بسيطين في اختيار العيّنة. لنُعد ترتيب الوحدات في الطبقة II بحيث يرد التركيب المرغوب ('5 ,'4 ,'3) أولاً مع 1 و 2 في الطبقة الأولى، كما نرى في الجانب الأيمن من الجدول (١٥ـ١٠). وعندئذ نسحب عددًا عشـوائيًا r بين 1 و 100 ونستخدمه لاختيار الوحدات من كل من الطبقتين. وفي الطبقة I، نختار الوحدة 1 إذا كان $1\leqslant r\leqslant 25$ ونختار الوحدة 2 إذا كان $26\leqslant r\leqslant 50$، وهكذا بحيث نعطي لكل وحدة الاحتمال المرغوب وهو $\frac{1}{4}$ في أن تكون هي الوحدة المختارة. وبصورة مماثلة، نختار '3 من الطبقة II إذا كان $1\leqslant r\leqslant 20$ ونختار '4 إذا كان $21\leqslant r\leqslant 40$ وهكذا. وبالتالي إذا كان $1\leqslant r\leqslant 20$ نختار ('1,3) وإذا كان $20\leqslant r\leqslant 25$ نختار ('1,4) وهلم جرًا. وتكون الاختيارات المشتركة واحتمالاتها كما يلي:

الزوج	(1, 3')	(1, 4')	(2, 4')	(2, 5')	(3', 5')	(3', 1)	(4', 1)	(4', 2)
الاحتمال	.20	.05	.15	.10	.10	.15	.05	.20

والتركيب الوحيد غير المرغوب هو ('3,5') وهكذا يكون الاحتمال الكلي للتراكيب المرغوبة 0.9.

وبما أن المعاينة ليست مستقلة في الطبقتين فلا تنطبق هنا العلاقات الخاصة بـ $V(\bar{y}_{st})$ و $v(\bar{y}_{st})$. ويعطي Riedel, Hess و Fitzpatrick (1976) علاقات تقريبية. وتعطي هذه المقالة أيضًا خوارزمية لتطبيق التحكم في الاختيار في مسائل بأكثر من طبقتين مع قيم لـ n وأساليب تحكّم أكثر تعقيدًا.

ومن أجل طريقة أخرى تستخدم تصاميم القطاعات غير التامة المتوازنة، انظر Avadhani و Sukhatme (1973).

(١٥ ـ ٧) بناء الطبقات

يثير هذا الموضوع عددًا من المسائل. ما هي الصفة المميزة الأفضل عند بناء الطبقات؟ كم يجب أن يوجد من الطبقات؟ ويبدو واضحًا في حالة مفردة واحدة أو

متغير واحد أن أفضل صفة مميزة نعتمدها للتقسيم إلى طبقات هو توزيع التكرار للمتغير y نفسه. وفي المحل الثاني ربما كان الشيء الأفضل هو توزيع التكرار لكمية أخرى تترابط مع المتغير بصورة عالية. وإذا علمنا عدد الطبقات فقد استنتج Dalenius (1957) المعادلات التي تقدم أفضل حدود للطبقات تحت المحاصتين التناسبية والنيمانية، كما قدم عدة باحثين طرقًا تقريبية أسرع. وسندرس المحاصّة النيمانية باعتبارها متفوقة على المحاصة التناسبية في مجتمعات تكون فيها المكاسب من التقسيم إلى طبقات كبيرة. ونفترض أولاً أن الطبقات مقامة باستخدام قيمة y نفسه.

لتكن y_0 و y_1 أصغر وأكبر قيم y في المجتمع، فالمسألة هي إيجاد الحدود الداخلية للطبقات $y_1, y_2, \ldots y_{L-1}$ بحيث يكون

$$V(\bar{y}_{st}) = \frac{1}{n}\left(\sum_{h=1}^{L} W_h S_h\right)^2 - \frac{1}{N}\sum_{h=1}^{L} W_h S_h^2 \qquad (5A.21)$$

أصغر ما يمكن. وإذا تجاهلنا الـ ت م م يكفي جعل $\sum W_h S_h$ أصغر ما يمكن. وبما أن y_h ظهر في هذا المجموع في الحدين $W_h S_h$ و $W_{h+1} S_{h+1}$ فقط فنجد:

$$\frac{\partial}{\partial y_h}(\sum W_h S_h) = \frac{\partial}{\partial y_h}(W_h S_h) + \frac{\partial}{\partial y_h}(W_{h+1} S_{h+1})$$

والآن إذا كان $f(y)$ دالة تكرار y فإن

$$W_h = \int_{y_{h-1}}^{y_h} f(t)\, dt, \qquad \frac{\partial W_h}{\partial y_h} = f(y_h) \qquad (5A.22)$$

بالإضافة إلى أن

$$W_h S_h^2 = \int_{y_{h-1}}^{y_h} t^2 f(t)\, dt - \frac{\left[\int_{y_{h-1}}^{y_h} t f(t)\, dt\right]^2}{\int_{y_{h-1}}^{y_h} f(t)\, dt} \qquad (5A.23)$$

وباشتقاق (5A.23) نجد

$$S_h^2 \frac{\partial W_h}{\partial y_h} + 2 W_h S_h \frac{\partial S_h}{\partial y_h} = y_h^2 f(y_h) - 2 y_h \mu_h f(y_h) + \mu_h^2 f(y_h) \qquad (5A.24)$$

حيث μ_h متوسط y في الطبقة h . وبإضافة $S_h^2\, \partial W_h/\partial y_h$ إلى الطرف الأيسر والكمية المساوية $S_h^2 f(y_h)$ إلى الطرف الأيمن نجد بعد التقسيم على $2S_h$

$$\frac{\partial(W_h S_h)}{\partial y_h} = S_h \frac{\partial W_h}{\partial y_h} + W_h \frac{\partial S_h}{\partial y_h} = \frac{1}{2} f(y_h) \frac{(y_h - \mu_h)^2 + S_h^2}{S_h} \qquad (5A.25)$$

وبصورة مماثلة نجد

$$\frac{\partial(W_{h+1} S_{h+1})}{\partial y_h} = -\frac{1}{2} f(y_h) \frac{(y_h - \mu_{h+1})^2 + S_{h+1}^2}{S_{h+1}} \qquad (5A.26)$$

وبالتالي تكون المعادلات الحسابية لِـ y_h هي

$$\frac{(y_h - \mu_h)^2 + S_h^2}{S_h} = \frac{(y_h - \mu_{h+1})^2 + S_{h+1}^2}{S_{h+1}} \quad (h = 1, 2, \ldots, L-1) \qquad (5A.27)$$

ومن سوء الحظ فإن هذه المعادلات لا تتكيف بسهولة مع الحسابات العملية، باعتبار أن كلًّا من μ_h و S_h يعتمد على y_h ونقدم هنا طريقة تقريبية سريعة تعود إلى Daleinius و Hodges (1959) كي نجعل $\sum W_h S_h$ أصغر ما يمكن. ليكن

$$Z(y) = \int_{y_0}^{y} \sqrt{f(t)}\, dt \qquad (5A.28)$$

وإذا كانت الطبقات ضيقة وكثيرة العدد، فينبغي أن يكون $f(y)$ ثابتًا تقريبًا (التوزيع المستطيل) ضمن طبقة معطاة. وبالتالي:

$$W_h = \int_{y_{h-1}}^{y_h} f(t)\, dt \doteq f_h (y_h - y_{h-1}) \qquad (5A.29)$$

$$S_h \doteq \frac{1}{\sqrt{12}} (y_h - y_{h-1}) \qquad (5A.30)$$

$$Z_h - Z_{h-1} = \int_{y_{h-1}}^{y_h} \sqrt{f(t)}\, dt \doteq \sqrt{f_h} (y_h - y_{h-1}) \qquad (5A.31)$$

حيث f_h القيمة «الثابتة» لِـ $f(y)$ في الطبقة h . وبتعويض هذه التقريبات نجد

$$\sqrt{12} \sum_{h=1}^{L} W_h S_h \doteq \sum_{h=1}^{L} f_h (y_h - y_{h-1})^2 \doteq \sum_{h=1}^{L} (Z_h - Z_{h-1})^2 \qquad (5A.32)$$

وبما أن $(Z_L - Z_O)$ مثبت فمن السهل التحقق من أن المجموع على اليمين يصبح أصغر بجعل $(Z_h - Z_{h-1})$ ثابتًا.

وفي حالة $f(y)$ معطى تكون القاعدة هي تشكيل القيم التراكمية لِـ $\sqrt{f(y)}$ واختيار الـ y_h بحيث تؤدي إلى فترات متساوية على سلّم القيم التراكمية لِـ $\sqrt{f(y)}$. ويوضح الجدول (١٥-١١) استخدام هذه القاعدة.

جدول (١٥-١١) حساب حدود الطبقات باستخدام قاعدة القيم المتجمعة لِـ $\sqrt{f(y)}$ Rule

القيم المتجمعة لِـ $\sqrt{f(y)}$	$f(y)$	% قروض صناعيّة / مجموع القروض	القيم المتجمعة لِـ $\sqrt{f(y)}$	$f(y)$	% قروض صناعية / مجموع القروض
340.3	126	50–55	58.9	3464	0–5
350.6	107	55–60	109.1	2516	5–10
359.7	82	60–65	155.5	2157	10–15
366.8	50	65–70	195.3	1581	15–20
373.0	39	70–75	229.1	1142	20–25
378.0	25	75–80	256.4	746	25–30
382.0	16	80–85	279.0	512	30–35
386.4	19	85–90	298.4	376	35–40
387.8	2	90–95	314.7	265	40–45
389.5	3	95–100	329.1	207	45–50

مثال

تبيّن المعلومات الإحصائية التوزيع التكراري للنسب المئوية من القروض المصرفية المخصصة لأغراض صناعية في مجتمع من 13435 مصرفًا في الولايات المتحدة، [1956, McEvory] . والتوزيع منحاز ومنواله يقع عند النهاية الدنيا. وفي عمود القيم المتجمعة لِـ $\sqrt{f}$ نجد $58.9 = \sqrt{3436},\ 109.1 = \sqrt{3464} + \sqrt{2516}$ ، وهكذا.

لنفرض أننا نريد خمس طبقات. بما أن المجموع أو القيمة المتجمعة الأخيرة لِـ $\sqrt{f}$ هي 389.5 فينبغي أن تكون نقاط التقسيم هي 77.9, 155.8, 233.7 و 311.6 على هذا السلّم. والنقاط المتوافرة الأكثر قربًا هي كما يلي:

	الطبقة				
	1	2	3	4	5
الحدود الفترة على	0–5%	5–15%	15–25%	25–45%	45–100%
القيم المتجمعة لِـ $\sqrt{f}$	58.9	96.6	73.6	85.6	74.8

والفترتان الأولى والثانية 58.9 و 96.6 مختلفتان كثيرًا، ولا يمكن تحسينهما دون مزيد من التقسيمات الجزئية للفئات الأصلية.

وإذا كانت فترات الفئات في التوزيع الأصلي لِـ y غير متساوية الطول، فإننا نحتاج إلى تغييرات طفيفة. وعندما تتغير الفترة من فترة طولها d إلى فترة طولها ud ، فإن قيمة $\sqrt{f}$ للفترة الثانية تُضرب عند تشكيل القيم المتجمعة لِـ $\sqrt{f}$ بِـ $\sqrt{u}$.

وقد اقترح Sethi (1963) طريقة أخرى لحساب الحدود المعطاة بالمعادلات الحسابية (5A.27) وذلك في حالة توزيع مستمر قياسي شبيه بالمجتمع الأصلي. وقد وضع Sethi جداول الحدود المثلى لمحاصة نيمانية أو متساوية أو تناسبية وذلك في حالة التوزيع الطبيعي والتوزيعات المختلفة وحيث يكون $L \leq 6$. وإذا ظهر أن أحد هذه التوزيعات يشكل تقريبًا للمجتمع موضع الدراسة فيمكن قراءة الحدود من جداول Sethi .

وتستدعي طريقتان تقريبيتان أُخْريان بعض التجربة والخطأ. فمن العلاقات (5A.27) نجد أن قاعدة Dalenius-Hodges مكافئة تقريبًا لجعل $W_h S_h$ ثابتًا، كما خمّن Dalenius و Gumey من قبل (1951) . وهناك قاعدة مماثلة لِـ Ekman (1959) وهي تجعل $W_h(y_h - y_{h-1})$ ثابتًا.

وبعد مقارنات تناولت بعض المجتمعات النظرية وثمانية مجتمعات للدراسة العملية وجد Cochran (1961) أن قاعدتي القيم التراكمية لِـ $\sqrt{f}$ وقاعدة Ekman قد أعطتا على الدوام نتائج جيدة (لم تُجرّب طريقة Sethi). وفي دراسة للسعة السريرية

لمستشفيات الولايات المتحدة والتي يشبه توزيعها التوزيع χ^2 بدرجة واحدة من الحرية، وجد Sethi, Hess و Balakrishnan (1966) أن طريقة Ekman تتفوق بصورة طفيفة على قاعدة القيم المتجمعة لـ $\sqrt{f}$ وطريقة Sethi في حالة $L>2$ ، بينما يفيد Murthy (1967) أيضًا بأداء جيد لطريقة Ekman .

وللعلاقات (5A.23) نتيجة مفيدة. فإذا كان W_hS_h ثابتًا تعطي محاصة نيمان حجم عيّنة ثابتا $n_h=n/L$ في جميع الطبقات. وفي الطرق التقريبية تقترح المقارنات التي تمّت أن القاعدة البسيطة $n_h=n/L$ هي قاعدة مُرضية.

وحتى الآن وضعنا فرضًا غير واقعي وهو أنه يمكن القيام بالتقسيم إلى طبقات على أساس قيم y نفسها. وعمليًا، يُستخدم متغير آخر x (ربما كان قيمة y في تعداد إحصائي حديث العهد). ويطوّر Murthy (1957) معادلات لحدود x التي تجعل $\sum W_hS_{yh}$ أصغر ما يمكن، علمًا بأننا نعلم انحدار y على x . وإذا كان هذا الانحدار غير خطي، فقد تختلف هذه الحدود اختلافًا كبيرًا عن الحدود المثلى التي نجدها عندما يكون المتغير x نفسه هو المتغير الذي سنقيسه. وتشير المعادلات، على أي حال، إلى أنه إذا كان انحدار y على x خطيًا والارتباط بين y و x مرتفعًا ضمن كل طبقة فينبغي أن تكون الحدود هي نفسها على وجه التقريب. ليكن

$$y=\alpha+\beta x+e$$

حيث $E(e)=0$ لجميع قيم x و x,e غير مترابطين. وتباين e ضمن الطبقة h هو $S_{eh}{}^2$. فعندئذ تحقق حدود المتغير x التي تجعل $V(\bar{y}_{st})$ أصغر ما يمكن المعادلات (1957, Dalinius) .

$$\frac{\beta^2[(x_h-\mu_{xh})^2+S_{xh}{}^2]+2S_{eh}{}^2}{\beta S_{xh}\sqrt{1+S_{eh}{}^2/\beta^2S_{xh}{}^2}}=\frac{\beta^2[(x_h-\mu_{x,h+1})^2+S^2_{x,h+1}]+2S^2_{e,h+1}}{\beta S_{x,h+1}\sqrt{1+S^2_{e,h+1}/\beta^2S^2_{x,h+1}}}$$

وإذا كان $S^2_{eh}/\beta^2S^2_{xh}$ صغيرًا من أجل جميع قيم h تُختصر هذه المعادلات إلى الشكل (5A.27) الذي يعطي الحدود المثلى لـ x ولكن $S^2_{eh}/\beta^2S^2_{xh}=(1-\rho_h{}^2)/\rho_h{}^2$ حيث ρ_h الارتباط ضمن الطبقة h بين x و y .

ومع أننا في حاجة إلى مزيد من التقصيّ إلا أن هذه النتيجة تقترح أن قاعدة القيم التراكمية لـ $\sqrt{f}$ مطبقة على x ينبغي لها أن تعطي تقسيمًا طبقيًا فعّالاً في حالة متغير آخر y انحداره على x خطي مع ارتباط مرتفع. وبعض النتائج العددية لـ Cochran (1961) تدعم هذا الحدس. وفضلاً عن ذلك، إذا كانت القيم معتدلة فقط، كما سيحدث عند زيادة عدد الطبقات، فإن العجز عن استخدام الحدود المثلى لـ x ينبغي أن يكون ذا قدر أقل من التأثير الضّار على y.

وبالطبع فإن المناقشة السابقة تلائم بصورة رئيسية معاينة مؤسسات قسمت إلى طبقات وفقًا لقياس ما للحجم. وتختلف الحالة عندما تكون مجموعة من المتغيرات على صلة وثيقة بالمتغير y_1 بينما تتصل مجموعة أخرى، توزيعها التكراري شديد الاختلاف عن توزيع المجموعة الأولى، اتصالاً وثيقًا بالمتغير y_2. وأحد الإمكانات هو التطلع إلى حل وسط بالنسبة لحدود الطبقات يواجه التساهلات المرغوبة في $V(\bar{y}_{1st})$ و $V(\bar{y}_{2st})$، متّبعين معالجة عامة أُعطيت في الفقرة (١٥ ـ ٤)، إلا أنه لا توجد بعد طرق حسابية لذلك.

وفي التقسيم الجغرافي إلى طبقات تكون المسألة أقل طواعية للمعالجة الرياضية باعتبار أن هناك العديد من الطرق المختلفة التي يمكن فيها تشكيل حدود الطبقات. والإجراء المعتاد هو اختيار متغيرات قليلة لها ارتباط عال مع المفردات الرئيسة في المسح، واستخدام مركّب من القوى الناتجة عن الخبرة ومن التجربة والخطأ لإقامة حدود جيّدة لهذه المتغيرات التي اخترناها. وبما أنه من المحتمل أن تكون المكاسب في الدقة الناشئة عن التقسيم إلى طبقات مكاسب متواضعة فليس جديرًا بالاهتمام أن نصرف قدرًا كبيرًا من الجهود في تحسين الحدود. وهناك أسس للتقسيم إلى طبقات في حالة مفردات اقتصادية ناقشها Stephan (1941)، Hagood و Bernert (1945) وفي حالة مفردات زراعية ناقشها King و Mc Carty (1941).

(١٥ ـ ٨) عدد الطبقات

والسؤالان المناسبان لقرار يتعلق بعدد الطبقات L هما: (ا) بأيّ معدّل يتناقص تباين $\bar{y}_{st}$ عند زيادة L ؟ (ب) كيف تتأثر تكلفة المسح الإحصائي بزيادة L ؟

وفيما يتعلق بـ (ا) لنفرض أن الطبقات قد أنشئت وفقًا لقيم y . ولأخذ الحالة الأكثر بساطة، لنفرض أن توزيع y هو التوزيع المستطيل في الفترة $(a,a+d)$. فعندئذ يكون S_y^2, قبل التقسيم إلى طبقات $d^2/12$ بحيث نجد $V(\bar{y})=d^2/12n$. في حالة عيّنة عشوائية بسيطة حجمها n ، وإذا أقمنا L من الطبقات ذات الحجوم المتساوية يكون التباين ضمن أي طبقة $S_{yh}^2=d^2/12L^2$. وبالتالي نجد في عيّنة طبقية فيها $W_h=1/L$ و $n_h=n/L$.

$$V(\bar{y}_{st})=\frac{1}{n}\left(\sum_{h=1}^{L} W_h S_{yh}\right)^2=\frac{1}{n}\left(\sum_{h=1}^{L}\frac{1}{L}\cdot\frac{d}{\sqrt{12}L}\right)^2=\frac{d^2}{12nL^2}=\frac{V(\bar{y})}{L^2} \quad (5A.33)$$

وهكذا يتناقص تباين $\bar{y}_{st}$ في حالة توزيع مستطيل عكسًا مع مربع عدد الطبقات، ومما يدهش أن هذه العلاقة تستمر في أن تبقى صحيحة بصورة تقريبية عندما نقسم إلى طبقات، توزيعات منحازة، واقعية، محدودة المدى، مع اختيار أمثل للحدود وفقًا لمحاصة نيمانية. وفي ثمانية توزيعات من النوع المحتمل وقوعه عمليًا، وجد Cochran (1961) أن القيم المتوسطة لـ $V(\bar{y}_{st})/V(\bar{y})$ كانت 0.232 ، 0.098 و 0.053 حيث $L=2,3,4$ وذلك بالمقارنة مع 0.250 ، 0.111 و 0.062 في حالة التوزيع المستطيل.

وهذه النتائج التي تقترح أن تكاثر الطبقات مفيد تعطي صورة مضللة لما يحدث عند استخدام متغير آخر x لتشكيل الطبقات. وإذا كان $\phi(x)=E(y/x)$ انحدار y على x ، فيمكن كتابة

$$y=\phi(x)+e \quad (5A.34)$$

حيث ϕ و e غير مرتبطين. وبالتالي،

$$S_y^2=S_\phi^2+S_e^2 \quad (5A.35)$$

ومن النتائج السابقة نعلم أن إنشاء L من الطبقات المثلى لـ x يمكن أن يخفّض S_ϕ^2

إلى S_ϕ^2/L^2 إذا كان $\phi(x)$ خطيًا، أو وفق نسبة أصغر إذا كان $\phi(x)$ غير خطي. إلا أن الحد S_e^2 لا ينخفض عند التقسيم إلى طبقات وفق x. ومع زيادة L سنصل إلى قيمة لِـ L يصبح معها S_e^2 الحد المسيطر. وسوف لا تُنتج أية زيادات أخرى في L إلا تخفيضًا نسبيًا لا يُذكر في $V(\bar{y}_{st})$. وتتوقف سرعة الوصول إلى مثل هذه النقطة على عدد من العوامل - وبصورة خاصة، على الحجمين النسبيين لِـ S_e^2 و S_ϕ^2 وطبيعة $\phi(x)$.

ولا يتـوافر في الأدبيات الإحصائية إلا أمثلة قليلة فقط من بيانات إحصائية واقعية. ولتعـويض هذا النقص نستخدم أسلوبًا نظريًا بسيطًا. فلنفرض أن الاختيار الأمثل لحدود الطبقة باستخدام x، مع عيّنات متساوية الحجوم n/L في كل طبقة، يخفّض $V(\bar{x}_{st})$ بمعدّل يتناسب مع $1/L^2$، فعندئذ يكون،

$$V(\bar{x}_{st}) = \frac{L}{n} \sum_{h=1}^{L} W_h^2 S_{xh}^2 = \frac{S_x^2}{nL^2} \qquad (5A.36)$$

لنفرض أيضًا أن انحدار y على x خطي، أي أن

$$y = \alpha + \beta x + e \qquad (5A.37)$$

حيث S_e^2 ثابت. وعندئذ

$$V(\bar{y}_{st}) = \frac{L}{n} \sum_{h=1}^{L} W_h^2 S_{yh}^2 = \frac{L\beta^2}{n} \sum_{h=1}^{L} W_h^2 S_{xh}^2 + \frac{LS_e^2}{n} \sum_{h=1}^{L} W_h^2 \qquad (5A.38)$$

وفي أي مجموعة من L من الطبقات لدينا $\sum W_h^2 \geq \frac{1}{L}$. وباستخدام (5A.36) نجد

$$V(\bar{y}_{st}) \geq \frac{1}{n}\left(\frac{\beta^2 S_x^2}{L^2} + S_e^2\right) = \frac{S_y^2}{n}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right] \qquad (5A.39)$$

حيث ρ الارتباط بين y و x في المجتمع غير المنقسم إلى طبقات.

ومع هذا النموذج يبيّن الجدول (٥ا-٢) $V(\bar{y}_{st})/V(\bar{y})$ في حالة ρ مساوٍ لِـ 0.99, 0.95, 0.90 و 0.85 و L يساوي 2 إلى 6، مفترضين أن العلاقة (5A.39) هي مساواة. وتعطي الأعمدة في الجانب الأيمن من الجدول قيمة $V(\bar{y}_{st})/V(\bar{y})$ لثلاث

مجموعات من المعلومات الإحصائية الواقعية، موصوفة تحت الجدول.

جدول (١٥ - ١٢) $V(\bar{y}_{st})/V(\bar{y})$ كدالة في L من أجل دالة انحدار خطي ومن أجل بعض البيانات الواقعية

	نموذج انحدار خطي				مجموعة بيانات		
	$\rho =$						
L	0.99	0.95	0.90	0.85	1	2	3
2	0.265	0.323	0.392	0.458	0.197	0.295	0.500
3	0.129	0.198	0.280	0.358	0.108	0.178	0.375
4	0.081	0.154	0.241	0.323	0.075	0.142	0.244
5	0.059	0.134	0.222	0.306	0.065	0.105	0.241
6	0.047	0.123	0.212	0.298	0.050	0.104	0.212
∞	0.020	0.098	0.190	0.277	—	—	—

نوع البيانات

مجموعة بيانات		x	y	المصدر
1	عدد المسجلين في الكلية	1952	1958	Cochran (1961)
2	مجموع المدن	1940	1950	Cochran (1961)
3	دخل الأسرة	1929	1933	Dalenius and Gurney (1951)

وتشير النتائج في حالة نموذج الانحدار الخطي إلى أنه ما لم يتجاوز ρ القيمـة 0.95 فينبغي أن لا نتـوقـع إلا القليل من الانخفـاض في التبـاين وراء $L=6$. والمجمـوعتـان 2 و 3 من البيانـات تدعـمان هذه النتيجـة، مع أن بعض الـزيادة الإضـافية في قيمـة L قد تكـون مفيدة في حالة بيانات التسجيل في الكلية (المجمـوعـة 1). وفي مقـارنتـين تتعلقـان ببيانـين إحصـائيين وجـد Sethi, Hess و Balakrishnan (1966) أن $V(\bar{y}_{st})$ يتناقص مع L بصورة أسرع مما تتنبأ به العلاقة، مما يشير إلى أن هناك مبالغة في تبسيط النموذج (5A.37) .

ولاستكمال هذا التحليل نحتاج إلى دالة تكلفة تبيّن كيفية اعتماد التكلفة على L ويقـترح Dalenius (1957) العـلاقة $C = LC_s + nC_n$. وستختلف نسبة التكلفتين C_s/C_n باختلاف نوع المسح الإحصائي. والزيادة في عدد الطبقات تتضمن عملاً إضافيًا في

تخطيط وسحب العيّنة كما تزيد عدد الترجيحات المستخدمة في حساب التقديرات ما لم تكن هذه ذاتية الترجيح . وفي بعض المسوح لا نحتاج تقريبًا الى أية تغيرات في تنظيم العمل الميداني؛ وفي مسوح أخرى نقيم وحدة ميدانية منفصلة في كل طبقة . ومهما كانت صيغة دالة التكلفة ، تقترح النتائج في الجدول (١٥ـ١٢) أنه إذا كانت زيادة L إلى أكثر من 6 ستضطرنا إلى أي تخفيض شديد في n ، كي نحفظ التكلفة ثابتة ، فنادرًا ما ستكون هذه الزيادة مفيدة .

وقد اقتصر النقاش في هذه الفقرة على مسوح نريد منها الوصول إلى تقديرات إجمالية . أما إذا أردنا أيضًا تقديرات من تقسيمات جغرافية جزئية للمجتمع فإن المنطق وراء عدد أكبر من الطبقات يُصبح أقوى .

(١٥ـ٩) التقسيم إلى طبقات بعد اختيار العيّنة
(التقسيم البَعدي الى طبقات)

مع بعض المتغيرات المناسبة لمسألة التقسيم إلى طبقات قد لا نعرف الطبقة التي تنتمي إليها الوحدة إلا بعد أن يتمّ جمع المعلومات الإحصائية . فالطبقات الشخصية مثل العمر، الجنس، العنصر، ومستوى التحصيل التربوي هي أمثلة تُذكر . وقد نتمكن من الحصول على حجوم الطبقات ، بدقة مقبولة ، من الإحصاءات الرسمية ، إلا أنه لا يمكن تصنيف الوحدات إلى الطبقات إلا بعد معرفة المعلومات الإحصائية من العيّنة . ونفترض هنا أن W_h و N_h معروفتان .

وإحدى الطرق هي أخذ عيّنة عشوائية بسيطة حجمها n ثم تصنيف الوحدات . وبدلاً من متوسط العيّنة $\bar{y}$ نستخدم التقدير $\bar{y}_w = \sum w_h \bar{y}_h$ حيث $\bar{y}_h$ متوسط وحدات المعاينة التي تقع في الطبقة h و $W_h = N_h/N$. ودقة هذه الطريقة هي تقريبًا دقة المعاينة العشوائية التناسبية نفسها شريطة أن يكون : (ا) حجم العيّنة كبير بصورة مقبولة ، ولنقُل أكبر من 20 ، في كل طبقة ، و(ب) من الممكن إهمال تأثيرات الأخطاء في الترجيحات W_h (انظر الفقرة ١٥ـ٢) .

ولتبيان ذلك ، ليكن m_h عدد الوحدات في العيّنة التي تقع ضمن الطبقة

h ، حيث تتغير m_h من عيّنة إلى عيّنة. ففي عيّنات تكون فيها الـ m_h مثبتة وجميعها تتجاوز الصفر نجد

$$V(\bar{y}_W) = \sum \frac{W_h^2 S_h^2}{m_h} - \frac{1}{N} \sum W_h S_h^2 \qquad (5A.40)$$

ويجب الآن حساب القيمة المتوسطة لِـ $V(\bar{y}_W)$ في عيّنات متكررة حجمها n وهذا يتطلب بعض الحذر، باعتبار أن واحدة أو أكثر من الـ m_h يمكن أن تكون صفرًا. وإذا حدث هذا، فسنضطر إلى ضمّ طبقتين أو أكثر قبل القيام بالتقدير، وقد نحصل عندئذ على تقدير أقل دقة. ومع ازدياد n يصبح احتمال كون أي من الـ m_h صفرًا احتمالاً صغيراً إلى حد تكون معه مساهمة هذا المصدر من مصادر التباين في القيمة النهائية للتباين مهملة.

وإذا تجاهلنا الحالة التي تكون فيها m_h صفرًا، فقد بيّن Stephan (1945) وإلى حدود من مرتبة n^{-2} أن

$$E\left(\frac{1}{m_h}\right) = \frac{1}{nW_h} + \frac{1 - W_h}{n^2 W_h^2} \qquad (5A.41)$$

وبالتالي

$$E[V(\bar{y}_W)] = \frac{1-f}{n} \sum W_h S_h^2 + \frac{1}{n^2} \sum (1 - W_h) S_h^2 \qquad (5A.42)$$

والحد الأول هو قيمة $V(\bar{y}_{st})$ في معاينة طبقية تناسبية. والحد الثاني يمثل الزيادة الناشئة عن عدم توزع المقادير m_h توزيعًا متناسبًا. إلا أن

$$\frac{1}{n^2} \sum (1 - W_h) S_h^2 = \frac{1}{n}\left(\frac{L}{n}\right) \bar{S}_h^2 - \frac{1}{n^2} \sum W_h S_h^2 = \frac{1}{n\bar{n}_h} \bar{S}_h^2 - \frac{1}{n^2} \sum W_h S_h^2 \qquad (5A.43)$$

حيث $\bar{S}_h^2$ هو معدل الـ S_h^2 و $\bar{n}_h = n/L$ هو العدد الوسطي للوحدات في الطبقة الواحدة. وهكذا إذا لم تختلف الـ S_h^2 اختلافًا كبيرًا فإن الزيادة تكون حوالي $(L-1)/L\bar{n}_h$ مرة التباين الموافق للمعاينة الطبقية التناسبية، متجاهلين هنا الـ ت م م. وستكون الزيادة صغيرة إذا كان $\bar{n}_h$ كبيرًا بصورة مناسبة.

ويمكن تطبيق هذه الطريقة أيضًا على عيّنة جرى تقسيمها إلى طبقات وفقًا لعامل آخر، مثلاً، إلى خمس مناطق جغرافية، شريطة أن تكون الـ W_h معروفة بصورة منفصلة ضمن كل منطقة. وهذا التقسيم إلى طبقات على مرحلتين مستخدم بصورة واسعة في المسوح الإحصائية القومية في الولايات المتحدة: انظر: Bean (1970) من أجل وصف لعلاقات التقدير في مسح المقابلة الصحية للمركز القومي للإحصاءات الصحية.

(١٥ ـ ١٠) المعاينة بالحصّة (الكوتا)

وفي طريقة أخرى استُخدمت في مسوح تتعلق بسبر الرأي وبحوث التسويق نحسب سلفًا الـ n_h التي نحتاجها في كل طبقة بحيث تكون المعاينة تناسبية. ونطلب من الإحصائي العدّاد أن يستمر في المعاينة حتى الحصول على الحصة الكوتا الضرورية في كل طبقة. والمتغيرات الأكثر استخدامًا في التقسيم إلى طبقات هي المنطقة الجغرافية، العمر، الجنس، العنصر، وبعض مقاييس السويّة الاقتصادية. وإذا كان العدّاد يختار الأشخاص عشوائيًا ثم يخصص كلاً منهم إلى طبقته المناسبة، فستكون الطريقة مطابقة لطريقة المعاينة العشوائية الطبقية. وقد نحتاج إلى قدر ضخم من العمل الميداني حتى نملأ جميع الحصص (الكوتا)، باعتبار أن معظم الأشخاص الذين نصل إليهم في المراحل المتأخرة قد يقعون في حصص تمّ ملؤها.

وللتعجيل في ملء الحصص يُخول العداد ببعض الحرية في العمل فيما يتعلق بالأشخاص أو المنازل التي سيختارها. ويتغير مقدار الحرية الممنوحة من وكالة إلى أخرى، إلا أنه يمكن، بصورة عامة وصف المعاينة بالحصة (الكوتا) بأنها معاينة طبقية مع اختيار غير عشوائي إلى حدّ ما للوحدات ضمن الطبقات. ولهذا السبب، لا يمكن الاطمئنان تمامًا إلى تطبيق العلاقات الخاصة بخطأ المعاينة على نتائج العينات بالحصة (الكوتا). وقد لخص Stephan و McCarty (1958) عددًا من المقارنات بين نتائج العيّنة بالحصة والعيّنة الاحتمالية، وأعطيا نقدًا ممتازًا لإنجاز كل من النوعين. ويبدو أنه من المحتمل أن تنتج طريقة الحصة (الكوتا) عيّنات منحازة بالنسبة لصفات مثل الدخل، التربية، والمهنة، مع أنها تتفق غالبًا بصورة جيدة مع العيّنات الاحتمالية في مسائل سبر الرأي والموقف.

(١٥-١١) التقدير من عيّنة للكسب العائد إلى التقسيم إلى طبقات

عند القيام بمعاينة عشوائية طبقية سيكون من المفيد، كدليل لتنفيذ مسوح إحصائية في المستقبل، أن نقدّر الكسب في الدقة بالنسبة إلى المعاينة العشوائية البسيطة.

والمعلومات الإحصائية المتوافرة من العيّنة هي قيم N_h ، n_h ، y_h وS_h^2 . ونرى من الفقرة (٥ - ٤) أن تقدير تباين المتوسط المرجّح من عيّنة طبقية هو

$$v(\bar{y}_{st}) = \sum \frac{W_h{}^2 s_h{}^2}{n_h} - \sum \frac{W_h s_h{}^2}{N}$$

وذلك وفقًا للعلاقة (5.13) .

والمسألة هي أن نقارن هذا التباين مع تقدير لتباين المتوسط الذي كان يمكن أن نحصل عليه من عيّنة عشوائية بسيطة. والطريقة المستخدمة أحيانًا هي أن نحسب المتوسط المعتاد لمربعات الانحرافات عن متوسط العيّنة

$$s^2 = \frac{\sum (y_{hi} - \bar{y})^2}{n-1}$$

حيث نتجاهل الطبقات. وتؤخذ هذه الكمية كتقدير لِـ s^2 بحيث إن $\hat{V}_{ran} = (N-n)s^2/Nn$ في حالـة متوسط عيّنة عشوائية بسيطة. وتعمل هذه الطريقة بجودة كافية إذا كانت المحاصة تناسبية، باعتبار أن العيّنة العشوائية البسيطة توزع نفسها بين الطبقات بصورة متناسبة تقريبًا. ولكن إذا تبنينا محاصة بعيدة عن كونها تناسبية فإن العيّنة المأخوذة فعلاً لا تشبه عيّنة عشوائية بسيطة، وهذه الكمية يمكن أن تكون تقديرًا رديئًا. ونعطي طريقة عامة يعود البرهان فيها إلى J.N.K. Rao (1962) .

نظرية (١٥ - ١)

بفرض نتائج عيّنة عشوائية طبقية يكون التقدير غير المنحاز لِـ V_{ran} ، تباين متوسط عيّنة عشوائية بسيطة من المجتمع نفسه هو:

(5A.44)

$$V_{ran} = \frac{(N-n)}{n(N-1)} \left[\frac{1}{N} \sum_{h}^{L} \frac{N_h}{n_h} \sum_{j}^{n_h} y_{hj}^2 - \bar{y}_{st}^2 + v(\bar{y}_{st}) \right]$$

حيث $v(\bar{y}_{st})$ هو التقدير غير المنحاز المعتاد لِـ $V(\bar{y}_{st})$.

برهان:

$$V_{ran}=\frac{(N-n)}{nN}S^2=\frac{(N-n)}{n(N-1)}\left[\frac{1}{N}\sum_h^L\sum_j^{N_h}y_{hj}^2-\bar{Y}^2\right] \quad (5A.45)$$

والآن

$$\frac{1}{N}E\left(\sum_h^L\frac{N_h}{n_h}\sum_j^{n_h}y_{hj}^2\right)=\frac{1}{N}\sum_h^L\sum_j^{N_h}y_{hj}^2 \quad (5A.46)$$

وأيضًا، وبما أن $v(\bar{y}_{st})$ و $\bar{y}_{St}$ تقديران غير منحازين لِـ $v(\bar{y}_{st})$ و $\bar{Y}$ على الترتيب

$$Ev(\bar{y}_{st})=V(\bar{y}_{st})=E(\bar{y}_{st}^2)-\bar{Y}^2 \quad (5A.47)$$

ومنه يكون التقدير غير المنحاز لِـ $\bar{Y}^2$ في (5A.45) هو

$$\bar{y}_{st}^2-v(\bar{y}_{st}) \quad (5A.48)$$

ومن (5A.46) و (5A.48) نجد أن تقدير عيّنة غير منحاز لِـ V_{ran} المذكور في (5A.45) هو

$$V_{ran}=\frac{(N-n)}{n(N-1)}\left[\frac{1}{N}\sum_h^L\frac{N_h}{n_h}\sum_j^{n_h}y_{hj}^2-\bar{y}_{st}^2+v(\bar{y}_{st})\right] \quad (5A.44)$$

وهذا يكمل برهان النظرية (١٥ ـ ١).

ومع محاصّة تناسبية $(N_h/n_h=N/n)$، يصبح الحدان الأولان داخل الأقواس المربعة مساويين لِـ$(\frac{1}{n})$ مرة مجموع المربعات ضمن العيّنة وهو $(n-1)s^2/n$ =. وتختزل العلاقة (5A.44) عندئذ إلى

$$v_{ran}=\frac{(N-n)}{n(N-1)}\left[\frac{(n-1)}{n}s^2+v(\bar{y}_{st})\right] \quad (5A.49)$$

وإذا كان n كبيرًا، $(N-1)\doteq N$، $(n-1)\doteq n$، فيكون الحد في $v(\bar{y}_{st})$ من مرتبة $\frac{1}{n}$ بالنسبة إلى الحد في s^2 المذكور في (5A.49) . وبالتالي نجد في حالة محاصّة تناسبية

$$v_{ran} \doteq \frac{(N-n)}{nN} s^2 \qquad (5A.50)$$

وفي الحالة العامة يكون التبسيط الموافق (n كبيرة) هو

$$v_{ran} \doteq \frac{(N-n)}{nN}\left[\frac{1}{N}\sum \frac{N_h}{n_h}\sum^{n_h} y_{hj}^2 - \bar{y}_{st}^2\right] \qquad (5A.51)$$

مثـــال:

سنوضح الحسابات من الطبقات الثلاث الأولى في العيّنة من كليات المعلمين (فقرة ٥ـ٩). والبيان الإحصائي في الجدول (١٥ـ١٣) يتعلق بالعيّنة المتأخرة عام 1946. وتمثل المتوسطات رقم التسجيل للكلية الواحدة بالآلاف [وقيم الـ s_h^2 أعلى بقليل مما في الطبقة الثانية]، وذلك يعود إلى التصحيح.

جدول (١٥ ـ ١٣) البيان الإحصائي الأساسي من عيّنة طبقية من كليات المعلمين

الطبقة	N_h	n_h	$\bar{y}_h$	s_h^2	$\frac{N_h}{n_h}\left(\sum^{n_h} y_{hj}^2\right)$
1	13	9	2.200	1.8173	83.920
2	18	7	1.638	0.0735	49.429
3	26	10	0.992	0.0859	27.596
	57	26			160.945

ومع n صغير نستخدم العلاقة (5A.44) فنجد $\bar{y}_{st}$=104715 ولم تتيسّر قيم الـ y_{hi} الخاصة بالعيّنة. إلا أنه يمكن الحصول على الأرقام في العمود الأيمن من الجدول (١٥ ـ ١٣) بالاستفادة من الأعمدة السابقة له. وبالتعويض في العلاقات المناسبة نجد

$$v(\bar{y}_{st}) = \frac{1}{N^2}\sum \frac{N_h(N_h-n_h)}{n_h} s_h^2 = 0.00497$$

$$v_{ran} = \frac{31}{(26)(56)}\left[\frac{160.945}{57} - (1.4715)^2 + 0.00497\right] = 0.01412$$

ويبدو أن التقسيم إلى طبقات قد خفّض التباين إلى حوالي ثلث قيمته في حالة عيّنة عشوائية بسيطة، وتقدير عامل الـ $deff$ (فقرة ٤ـ١١) هو 0.35 = 0.00497/ 0.01412 .

(١٥ - ١٢) تقدير التباين
في حالة وحدة معاينة واحدة في كل طبقة

إذا كان المجتمـع متغـيراً بصورة مرتفعة، ونعرف العديد من القواعد الفعالة للتقسيم إلى طبقات، فيمكن المضيّ في التقسيم إلى النقطة التي تحوي معها العينة وحدة واحدة فقط في كل طبقة. وفي هذه الحالة لا يمكن استخدام العلاقات المعطاة سابقًا لتقدير $V(\bar{y}_{st})$ و $V(\hat{Y}_{st})$ ومع L زوجي يمكننا محاولة القيام بتقدير من خلال تجميع الطبقات في أزواج يُعتقد مقدمًا أن مجموعهما متساويان تقريبًا. وينبغي القيام بالتجميع في أزواج قبل رؤية نتائج العيّنة، وذلك لأسباب ستتضح فيما بعد.

لتكن ملاحظات العيّنة في زوج نموذجي y_{j1} ، y_{j2} حيث يتحول j من 1 الى $L/2$. وليكن $\hat{Y}_{j1}=N_{j1}y_{j1},\ \hat{Y}_{j2}=N_{j2}y_{j2}$ تقديري مجموعي الطبقتين. فنجد الآن،

$$\hat{Y}_{j1}-\hat{Y}_{j2}=(Y_{j1}-Y_{j2})+(\hat{Y}_{j1}-Y_{j1})-(\hat{Y}_{j2}-Y_{j2}) \qquad (5A.52)$$

وبأخذ المتوسط فوق جميع العيّنات من هذا الزوج نجد

$$E(\hat{Y}_{j1}-\hat{Y}_{j2})^2=(Y_{j1}-Y_{j2})^2+N_{j1}(N_{j1}-1)S_{j1}^2+N_{j2}(N_{j2}-1)S_{j2}^2 \qquad (5A.53)$$

ومن أجل $V(\hat{Y}_{st})$ لنأخذ التقدير،

$$v_1(\hat{Y}_{st})=\sum_{j=1}^{L/2}(\hat{Y}_{j1}-\hat{Y}_{j2})^2 \qquad (5A.54)$$

ومن (5A.53) تكون القيمة المتوقعة لهذا المقدار،

$$Ev_1(\hat{Y}_{st})=\sum_{h=1}^{L}N_h(N_h-1){S_h}^2+\sum_{j=1}^{L/2}(Y_{j1}-Y_{j2})^2 \qquad (5A.55)$$

والحد الأول على اليمين هو التباين الصحيح (استنادًا إلى النظرية ٥-٤ مع $n_h=1$) ويمثل الحد الثاني انحيازًا موجبًا، يعتمد حجمه على النجاح الذي نبلغه في اختيار أزواج من الطبقات تختلف مجاميعها الصحيحة اختلافًا بسيطًا بعضها عن بعض. ويحذّر شكل التقدير في (5A.54) من أن تشكيل الأزواج، من خلال جعل الاختلاف

في المجاميع المقدرة من العيّنة اختلافًا بسيطًا قدر الإمكان، يمكن أن يؤدي إلى تخفيض جديد في التقدير. وتدعى الطريقة طريقة «الطبقات المنهارة».

وفي حالة L فردي فيجب بالطبع أن توجد زمرة واحدة حجمها مختلف عن 2. وتعميم التقدير (5A.45) إلى G من الزمر من أية حجوم نختارها $L_j \geq 2$ يؤدي إلى

$$v_1(\hat{Y}_{st}) = \sum_{j=1}^{G} \frac{L_j}{L_j - 1} \sum_{k=1}^{L_j} (\hat{Y}_{jk} - \hat{Y}_j/L_j)^2 \qquad (5A.56)$$

حيث $\hat{Y}_j$ المجموع المقدّر للزمرة j وفي حالة $L_j = 2$، (وعندها يكون $\hat{Y}_j = \hat{Y}_{j1} + \hat{Y}_{j2}$، يتفق هذا الشكل مع (5A.54) وكما في (5A.54) يعطي توقع هذا الـ $v_1(\hat{Y}_{st})$ التباين الصحيح $V(\hat{Y}_{st})$ بالإضافة إلى انحياز موجب نجده بتعويض Y_j و Y_{jk} بدلاً من $\hat{Y}_j$ و $\hat{Y}_{jk}$ في (5A.56).

وإذا كنا نعلم في كل طبقة متغيراً مساعداً A_h نتنبأ بوساطته بمجموع الطبقة y_h فيقترح Hansen, Hurwitz و Madow (1953) تقديرًا بديلاً للتباين هو:

$$v_2(\hat{Y}_{st}) = \sum_{j=1}^{G} \frac{L_j}{L_j - 1} \sum_{k=1}^{L_j} (\hat{Y}_{jk} - A_{jk}\hat{Y}_j/A_j)^2 \qquad (5A.57)$$

وإذا كان A_h وسيلة تنبؤ جيدة، فمن المحتمل أن يكون هذا الانحياز الموجب في عبارة v_2 الناشيء عن الانحرافات $(\hat{Y}_{jk} - A_{jk}\hat{Y}_j/A_j)^2$ أصغر من الحد الموافق في عبارة v_1 مع أن v_2، وعلى العكس من v_1، يعطي أيضًا تقديرًا منحازًا للحد الذي يتضمن S_h^2 في عبارة $V(\hat{Y}_{st})$. وقد وجد Hartley, Rao و G. Kiefer (1969) أن v_2 أقل انحيازًا من v_1 في مجتمعين من ثلاثة مجتمعات، ووجدوا فرقًا بسيطًا في المجتمع الثالث.

وقد طوّر هؤلاء الباحثون طريقة لا تتضمن إنهيار الطبقات. وهذه الطريقة تستخدم واحدًا أو أكثر من المتغيرات المساعدة x_{1h}، x_{2h}، وهلمّ جرًا، التي يُعتقد أن للمتوسطات الحقيقية للطبقات $\bar{Y}_h$ انحدارًا خطيًّا عليها. وإذا كانت y_h قيمة العيّنة في الطبقة h فالطريقة تستخدم الانحرافات:

$$d_h = y_h - \bar{y} - \sum_i b_i(x_{ih} - \bar{x}_i) \qquad (5A.58)$$

ويمكن التعبير عن مصفوفة التباين للمقادير d_h كدالة خطية في المقادير σ_h^2 بالإضافة إلى حدود انحياز معينة . ومن هذه العلاقة نحصل على تقديرات $\hat{\sigma}_h{}^2$ حيث

$$\sigma_h{}^2 = (N_h - 1)S_h{}^2/N_h,$$

مما يعطي

$$v_3(\hat{Y}_{st}) = \sum_{h=1}^{L} N_h{}^2 \hat{\sigma}_h{}^2 \tag{5A.59}$$

ويبدو أن الطريقة تدعو للتفاؤل ويمكن تمديدها إلى التقديرات النسبة ، إلا أن الباحثين يحذرون من أنه ينبغي القيام بمقارنات إضافية مع طرق الطبقات المنهارة . وباستخدام أسلوب مختلف في المعالجة ، طوّر Fuller (1970) طريقة لإنشاء الطبقات تعطي تقدير عيّنة غير منحاز لِـ $V(\hat{Y}_{st})$ مع وحدة معاينة واحدة لكل طبقة . وتوخيًا للبساطة لنفرض أن $N/n = N/L = k$ (k عدد صحيح) ولنختر عددًا عشوائيًا r بين 1 و k فالطبقة الأولى تتألف من الوحدات المرقمة من $r+1$ إلى $r+k$ والثانية تلك المرقمة من $r+k+1$ إلى $(r+2k)$ وهكذا ، والأخيرتان (الطبقة الـ L والطبقة الـ n) هما المرقمتان من $r+(n-1)k+1$ إلى $N=nk$ وتلك المرقمة من 1 إلى r . وللوهلة الأولى ، قد تبدو هذه الطبقة الأخيرة اختيارًا غير موفّق . وكما يلاحظ Fuller على أي حال، يمكن لهذه الطريقة أن تعمل جيدًا في التقسيم الجغرافي إلى طبقات مع وحدات معاينة على شكل مساحة . وهنا يعتمد التقسيم إلى طبقات ، عادة ، على فكرة أن الوحدات القريبة بعضها من بعض تميل إلى أن تكون متشابهة . وبترقيم الوحدات بطريقة أفعوانية يمكن أن نجد y_N إلى جوار y_1 ، بحيث تكون الطبقة المتضمنة لكل من y_N و y_1 متجانسة داخليًا أيضًا . فالتقدير $v(\hat{Y}_{st})$ هو مجموع مرجّح للفروق $(y_h - y_{h+1})^2$.

وقد تكون الطريقة الدائرية أقل فعالية في مجتمع يُظهر نزعة إلى التزايد من y_1 إلى y_N وعندئذ سيكون للطبقة التي تتضمن y_1 و y_N تغير داخلي كبير . وفي هذه الحالة يعطي Fuller خطة ثانية أكثر تعقيدًا بقليل، وينبغي لها أن تعطي دقة جيدة، في حالة وجود نزعة إلى التزايد ، كما تزودنا أيضًا بتقدير غير منحاز لِـ $V(\hat{Y}_{st})$.

(١٥ ـ ١٣) الطبقات بصفتها ميادين دراسة

تعالج هذه الفقرة مسوحًا إحصائية تكون غايتها الرئيسة القيام بمقارنات بين طبقات مختلفة نفترض إمكانية تحديدها سلفًا. وتختلف قواعد تحديد حجوم العيّنات من الطبقات المختلفة عن تلك التي نطبقها عندما يكون الهدف هو القيام بتقديرات إجمالية تتعلق بالمجتمع. وإذا كان لدينا طبقتان فقط فيمكن اختيار n_1 و n_2 بحيث نجعل تباين الفرق $(\bar{y}_1-\bar{y}_2)$ بين تقديري متوسطي الطبقتين أصغر ما يمكن. وبحذف الـ ت م م لأسباب أُعطيت في الفقرة (٢ ـ ١٤)، نجد

$$V(\bar{y}_1-\bar{y}_2)=\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2} \qquad (5A.60)$$

ومع دالة تكلفة خطية كالدالة

$$C=c_0+c_1n_1+c_2n_2 \qquad (5A.61)$$

يكون V أصغر ما يمكن عندما يكون

$$n_1=\frac{\dfrac{nS_1}{\sqrt{c_1}}}{S_1/\sqrt{c_1}+S_2/\sqrt{c_2}}, \qquad n_2=\frac{\dfrac{nS_2}{\sqrt{c_2}}}{S_1/\sqrt{c_1}+S_2/\sqrt{c_2}} \qquad (5A.62)$$

وفي حالـة L من الطبقات، $L>2$ تعتمد المحاصة المثلى على مقدار الدقة المرغوبة في مقارنات مختلفة. وعلى سبيل المثال، يمكن جعل التكلفة أصغر ما يمكن، خاضعة لمجموعة من $L(L-1)/2$ من الشروط المتمثلة في $V(\bar{y}_h-\bar{y}_i)\leq V_{hi}$، حيث نختار القيم V_{hi} وفقـا للدقة التي نرى ضرورتها لتحقيق مقارنة مُرضية بين الطبقتين h و i. وكثيرًا ما نجد أن طريقة أبسط في المحاصّة هي طريقة مناسبة، خاصة إذا لم تختلف المقادير S_h والمقادير C_h كثيرًا. وأحد الأساليب هو جعل التباين المتوسط للفرق بين الـ $L(L-1)/2$ من أزواج الطبقات أصغر ما يمكن، أي أخذ النهاية الصغرى لِـ

$$\bar{V}=\frac{2}{L}\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}+\cdots+\frac{S_L^2}{n_L}\right) \qquad (5A.63)$$

ونجعل $\bar{V}$ أصغر ما يمكن، في حالة C مثبتة، باستخدام القاعدة المذكورة في (5A.62)،

$$n_h \propto \frac{S_h}{\sqrt{c_h}} \qquad (5A.64)$$

ويمكن أن تُنتج هذه القاعدة أزواجًا معينة من الطبقات تُقارن بدقة أكبر بينما تُقارن أزواج أخرى بدقة أقل مما نشعر أنه الدقة المناسبة. وإحدى الطرق البديلة هي اختيار الـ n_h بحيث يبقى الخطأ المعياري $\sqrt{V}$, مثلاً، نفسه من أجل كل زوج من الطبقات. وهذا يؤدي إلى جعل $S_h^2/n_h = V/2$ في كل طبقة. وفي حالة تكلفة مثبتة تعطي هذه الطريقة دقة إجمالية أقل مما تعطيه الطريقة الأولى. ويمكن للقارىء أن يتحقق من أن المحاصّتين المُثليين تعطيان

$$\bar{V} = \frac{2(\sum S_h\sqrt{c_h})^2}{L(C-c_0)}, \qquad V = \frac{2(\sum S_h^2 c_h)}{(C-c_0)} \qquad (5A.65)$$

ونستنتج من متراجحة كوشي - شوارتز أن V أكبر من $\bar{V}$ على الدوام ما لم يكن $S_h\sqrt{c_h}$ ثابتًا. وإذا كان V أكبر بكثير من $\bar{V}$ فيمكن أحيانًا، وبعد قليل من التجربة والخطأ، الوصول إلى محاصة تشكل حلًا وسطًا، وتعطي تباينًا وسطيًا قريبًا من $\bar{V}$ ، وتُبقي $V(\bar{y}_h - \bar{y}_i)$ ثابتًا إلى حد ما.

ويكون الهدف أحيانًا الحصول على تقديرات لكل طبقة بالإضافة إلى تقديرات إجمالية من أجل المجتمع ككل. وعند تخطيط المسح الإحصائي يمكن تحديد الشروط التالية

$$V(\bar{y}_h) = \frac{S_h^2}{n_h}(1-f_h) \le V_h, \qquad V(\bar{y}_{st}) = \sum \frac{N_h^2 S_h^2}{n_h}(1-f_h) \le V$$

وحدود الـ ت م م مأخوذة هنا في الاعتبار مادامت الغاية هي تحديد دقة تقدير المتوسطات التي نريد تقديرها في مجتمع منته. وتحدد الشروط المتعلقة بـ $V(\bar{y}_h)$ حدودًا دنيا لقيم الـ n_h . وإذا وجدنا أن هذه الحدود الدنيا تحقق الشروط الخاصة بـ $V(\bar{y}_{st})$, تكون مسألة المحاصة قد حُلت. وعندما لا يتحقق الشرط المتعلق بـ $V(\bar{y}_{st})$ يشير Dalenius (1975) إلى طريقة بيانية.

وتنشأ مسائل أكثر تعقيدًا عندما تمثل الـ $L=2^k$ من الطبقات جميع تراكيب k من العوامل، ولكل عامل مستويان، أما الهدف فهو تقدير متوسطات تأثيرات العوامل. وإذا كانت الطبقة أو الخلية التي ينتمي إليها أي عنصر من المجتمع معروفة قبل المعاينة، فيمكن سحب عيّنة بالحجم المرغوب n_h من الطبقة h . وفي حالة 3,2 أو 4 عوامل، أعطى Sedransk (1967) طرقًا لإيجاد الـ n_h التي تجعل التكلفة أصغر ما يمكن تحت مواصفات مختلفة تتعلق بتباينات متوسطات التأثيرات المقدّرة للعوامل، كما تتعلق بالقوة المرغوبة لاختبار حول التفاعلات.

(١٥ ـ ١٤) تقديـر المجاميـع والمتوسطات فوق مجتمعات جزئية

كثيرًا ما تكون المجتمعات الجزئية أو ميادين الدراسة ممثّلة في جميع الطبقات. وإذا كان التقسيم إلى طبقـات جغرافيًا مثلاً، فقد نرغب في تقديرات منفصلة فوق المجتمع ككل، للذكور والإناث أو لفئات الأعمار المختلفة أو لمستخدمي معجون للأسنان ومن لا يستخدمونه، وما شابه. وتقدم المسألة بعض التعقيدات. وقد أعطى Yates (1953) العلاقات الأساسية وأضاف Durbin (1958) و Hartley (1959) مزيدًا من المناقشة والبراهين. كما نوقشت الطرق القابلة للتطبيق على طبقة واحدة في الفقرتين (٢ـ١٠) و(٢ـ١١). وتنطبق الرموز التالية على وحدات من طبقة h واقعة في الميدان j .

رمـوز

عدد الوحدات: $N_{hj}, \quad \sum_j N_{hj} = N_h$

عدد الحدات في العيّنة: $n_{hj}, \quad \sum_j n_{hj} = n_h$

القياس من وحدة بمفردها: y_{hij}

متوسط العيّنة: $\bar{y}_{hj} = \sum_{i=1}^{n_{hj}} \frac{y_{hij}}{n_{hj}}$

متوسط الميدان : $\bar{Y}_{hj} = \sum_{i=1}^{N_{hj}} \frac{y_{hij}}{N_{hj}}$

ومجموع ومتوسط المجتمع في الميدان j فوق جميع الطبقات، هما على الترتيب،

$$Y_j = \sum_h N_{hj} \bar{Y}_{hj}, \qquad \bar{Y}_j = \frac{Y_j}{N_j}$$

حيث $N_j = \sum_h N_{hj}$.

وتنشأ التعقيدات بسبب كون الـ n_{hj} متغيرات عشوائية. وإذا كانت الـ N_{hj} معروفة، فستكون المسألة بسيطة. وكتقديرين لـ Y_j و $\bar{Y}_j$ يمكننا استخدام

$$\hat{Y}_h{}' = \sum_h N_{hj} \bar{y}_{hj}, \qquad \hat{\bar{Y}}_j{}' = \frac{\hat{Y}_j{}'}{N_j}$$

واستنادًا إلى الطريقة المذكورة في الفقرة (٢-١٢)، لا تزال العلاقة العادية الخاصة بـ $V(\bar{y}_{hj})$ صحيحة، شريطة أن تكون جميع الـ n_{hj} موجبة. وهكذا نجد،

$$V(\hat{Y}_j{}') = \sum_h \frac{N_{hj}^2 S_{hj}^2}{n_{hj}} \left(1 - \frac{n_{hj}}{N_{hj}}\right) \qquad (5A.66)$$

حيث S_{hj}^2 التباين بين الوحدات في الميدان j ضمن الطبقة h . وعلى أي حال، نادرًا ما تكون الـ N_{hj} معروفة.

تقدير مجاميع الميادين

في حال عدم معرفة الـ N_{hj} نقدر مجموع كل طبقة من الميدان كما في الفقرة (٢-١٣). وتضاف هذه المجاميع لنحصل على تقدير لمجموع الميدان، أي أن،

$$\hat{Y}_j = \sum_h \frac{N_h}{n_h} \sum_i^{n_{hj}} y_{hij} \qquad (5A.67)$$

ونجد التباين الصحيح والمقدَّر لـ $\hat{Y}_j$ وفقًا للتدابير المستخدمة في الفقرة (٢-١٣). ونُدخل الآن متغيرًا y'_{hi} يساوي y_{hij} في جميع وحدات الميدان j ويساوي

الصفر في جميع الوحدات الأخرى في المجتمع. وكما بيّنا في الفقرة (٢-١٣) يعطي هذا من أجل التباين المقدّر،

$$v(\hat{Y}_j) = \sum_h \frac{N_h^2}{n_h(n_h-1)}(1-f_h)\left[\sum_i^{n_{hi}} y_{hij}^2 - \frac{(\sum y_{hij})^2}{n_h}\right] \tag{5A.68}$$

تقدير متوسطات الميادين

لكي نقدّر متوسط ميدان Y_j/N_j نحتاج إلى تقدير عيّنة لـ N_j. وكتقدير غير منحاز نجد

$$\hat{N}_j = \sum_h \frac{N_h}{n_h} n_{hj} \tag{5A.69}$$

وبالتالي نأخذ،

$$\hat{\bar{Y}}_j = \frac{\hat{Y}_j}{\hat{N}_j} = \frac{\sum_h (N_h/n_h) \sum_i y_{hij}}{\sum_h (N_h/n_h) n_{hj}} \tag{5A.70}$$

وفي حالة تقسيم متناسب إلى طبقات يصبح $\hat{\bar{Y}}_j$ متوسط العيّنة العادي للوحدات الواقعة في الميدان j. وفي الحالة العامة، يُعرف هذا التقدير بأنه التقدير النّسبة المركّب، الذي سنناقشه فيما بعد في الفقرة (٦-١١). ولتبيان ذلك، نُدخل متغيراً مصطنعًا آخر يساوي 1 من أجل كل وحدة من الميدان j وصفرًا من أجل جميع الوحدات الأخرى، وحيث يتغير i من 1 إلى N_h. ومن الواضح أن

$$\bar{x}_h' = \frac{\sum_i^{n_h} x_{hi}'}{n_h} = \frac{n_{hj}}{n_h}, \qquad \bar{y}_h' = \frac{\sum_i^{n_h} y_{hi}}{n_h} = \frac{\sum_i^{n_{hj}} y_{hij}}{n_h} = \frac{n_{hj}}{n_h}\bar{y}_{hj} \tag{5A.71}$$

بحيث يمكن كتابة تقدير متوسط الميدان على الشكل،

$$\hat{\bar{Y}}_j = \frac{\sum_h (N_h/n_h) \sum_i y_{hij}}{\sum_h (N_h/n_h) n_{hj}} = \frac{\sum_h N_h \bar{y}_h'}{\sum_h N_h \bar{x}_h'} = \frac{\bar{y}_{st}'}{\bar{x}_{st}'} \tag{5A.72}$$

وهي العلاقة الخاصة بالتقدير النسبة المركّب للمتغيرين y_{hi}' و x_{hi}'. ومن الفقرة (٦-١١) يمكن التعبير، بصورة تقريبية، عن تقدير التباين على الشكل،

$$v(\hat{\bar{Y}}_j) \doteq \frac{1}{\hat{N}_j^2} \sum_h \frac{N_h^2(1-f_h)}{n_h(n_h-1)} \sum_i^{n_h} [y_{hi}' - \hat{\bar{Y}}_j x_{hi}' - (\bar{y}_h' - \hat{\bar{Y}}_j \bar{x}_h')]^2 \qquad (5A.73)$$

وباستخدام (5A.71) يمكن كتابة المجموع الثاني على الشكل،

$$\sum_i^{n_h} (y_{hi}' - \hat{\bar{Y}}_j x_{hi}')^2 - n_h(\bar{y}_h' - \hat{\bar{Y}}_j \bar{x}_h')^2 = \sum_i^{n_{hj}} (y_{hij} - \hat{\bar{Y}}_j)^2 - \frac{n_{hj}^2}{n_h}(\bar{y}_{hj} - \hat{\bar{Y}}_j)^2 \qquad (5A.74)$$

وفضلاً عن ذلك، يمكن التعبير عن الحد الأول من (5A.74) بصورة بديلة على الشكل:

$$\sum_i^{n_{hj}} (y_{hij} - \bar{y}_{hj})^2 + n_{hj}(\bar{y}_{hj} - \hat{\bar{Y}}_j)^2$$

وبتعويض هذه النتائج في (5A.75) نجد أخيراً من أجل تقدير التباين،

$$v(\hat{\bar{Y}}_j) \doteq \frac{1}{\hat{N}_j^2} \sum_h \frac{N_h^2(1-f_h)}{n_h(n_h-1)} \left[\sum_i (y_{hij} - \bar{y}_{hj})^2 + n_{hj}\left(1 - \frac{n_{hj}}{n_h}\right)(\bar{y}_{hj} - \hat{\bar{Y}}_j)^2\right] \qquad (5A.75)$$

ويمثل الحد الأيمن مساهمة ما بين الطبقات في التباين. ولم نحذف الفروق بين متوسطات الطبقات، حذفاً كاملاً، من تباين المتوسط المقدّر لأي مجتمع جزئي. وتكون مساهمة ما بين الطبقات صغيرة إذا كانت الحدود $1 - n_{hj}/n_h$ صغيرة، أي إذا كان المجتمع الجزئي في مثل حجم المجتمع بكامله تقريبًا.

وكما أشار Durbin (1958) تنطبق (5A.75) أيضًا على متوسطات قدّرناها للمجتمع بكامله، إذا كانت العيّنة غير كاملة لأي سبب من الأسباب مثل عدم الاستجابة، شريطة أن يكون $\hat{\bar{Y}}_i$ هو بالطبع التقدير المستخدم. ويوجد، على أي حال، تعقيد إضافي يتمثل في أن للجزء «غير المستجيب» من المجتمع متوسطًا يختلف في الغالب عن متوسط الجزء «المستجيب». وهكذا يكون $\hat{\bar{Y}}_i$ تقديرًا منحازًا لمتوسط المجتمع بكامله، و (5A.75) لا تتضمن مساهمة هذا الانحياز.

(١٥ ـ ١٥) المعاينة من إطارين

في المسح الإحصائي لمخازن التجزئة الذي قام به مكتب الإحصاء عام 1949 والموصوف في Hurwitz, Hansen و Madow (p. 516 ff., 1953) نجد مثالاً قديمًا عن استخدام عيّنة من قائمة B من محلات الأعمال الكبيرة مع عيّنة من إطار مساحي A يغطي كامل المجتمع. والأهداف من هذا الاستخدام المركب لإطار قائمة غير كاملة مع إطار مساحي كامل هو كسب المزيد من الدقة وتوفير المال. ويمكن معاينة المحلات التجارية الكبيرة أحيانًا بتكلفة رخيصة بواسطة البريد والهاتف معًا؛ وفضلاً عن ذلك، وباعتبارها كبيرة، فهي في الغالب المحلات التي تتصف بأكبر تباين بالنسبة للمتغيرات y المقاسة. وبوضعها في طبقة مستقلة مع محاصة مثلى (أو معاينة نسبتها 100% إذا بدا ذلك قريبًا من الوضع الأمثل) يمكن أن نحصل على زيادة كبيرة في الدقة. وفي مسح مخازن التجزئة حُددت المحلات من إطار القائمة التي توجد في العيّنة من المساحات ثم أُخرجت من العيّنة المساحية، بحيث قُسم المجتمع موضع المعاينة إلى طبقتين متميزتين. وقد أُنجزت هذه العملية، وتدعى العملية الازدواجية، من قبل المشرف الميداني قبل المعاينة. وأحيانًا تتعقد الازدواجية وتخضع للأخطاء. وقد ناقش Hurwitz, Hansen و Jabine (1963) بعض الصعوبات التطبيقية بالإضافة إلى طرق مختلفة يمكن فيها للقوائم غير الكاملة أن تكون ذات فائدة.

وتحديد العناصر من عيّنة الإطار A التي تنتمي إلى إطار القائمة B يتطلّب أحيانًا قياس المتغيرات y من أجل هذه العناصر. ويكون للمعاين عندئذ ثلاث عيّنات تحت تصرفه وقد جرى فيها جميعًا قياس y : عيّنة من الطبقة a (الجزء من A الذي لا ينتمي إلى B) وعيّنتان مستقلتان من الطبقة B . إحداهما نرمز لها بالدليل ab وهي العيّنة التي نحصل عليها من A وعلمنا أنها تنتمي إلى B ، والأخرى نحصل عليها بمعاينة مباشرة من الإطار B . وفي حالة كمال الإطار A ومعاينة عشوائية بسيطة من كل من الإطارين، يقترح Hartley (1962) استخدام كلتا العيّنتين من B في تقديرات ما بعد التقسيم إلى طبقات

$$\hat{Y} = N_a\bar{y}_a + N_{ab}(p\bar{y}_{ab} + q\bar{y}_B) \tag{5A.76}$$

حيث $\bar{y}_a, \bar{y}_{ab}, \bar{y}_B$ ترمز إلى متوسطات العيّنات على الترتيب. وقد اختير عاملا الترجيح p و q من أجل العيّنتين اللتين تنتميان إلى B ، حيث $p+q=1$ كي يجعلا $V(\hat{Y})$ أصغر ما يمكن تحت دالة تكلفة من الشكل،

$$C = c_A n_A + c_B n_B \tag{5A.77}$$

وفي (5A.76) ستكون حجوم الطبقات و$N_{ab} = N_B$ و $N_a = (N_A - N_B)$ بالطبع معروفة.

ومع $c_B < c_A$، و $S_B{}^2 > S_a{}^2$ ، بيّن Hartley أن هذه الطريقة يمكن أن تؤدي إلى تخفيض كبير في $V(\hat{Y})$ بالمقارنة مع المعاينة من الإطار A لوحده، وذلك حتى عندما نقسم العيّنة من الإطار A إلى الطبقتين a و $B=ab$.

وتصبح المسألة أكثر صعوبة إذا كان الإطار A غير تام أيضًا. حيث نحتاج إلى إطارين A و B ، مع بعض الازدواج، للحصول على تغطية تامة للمجتمع. ومن أجل ما بعد التقسيم إلى طبقات توجد ثلاث طبقات متميزة: a (وحدات في A بمفردها)؛ ab (وحدات في كل من A و B) ؛ و b (وحدات في B بمفردها). ولا يمكن معاينة الطبقات الثلاث بصورة مباشرة، ويجب سحب عيّنات حجومها n_A, n_B من الإطارين A و B . وفضلاً عن ذلك، سوف لا تكون حجوم الطبقات N_a, N_{ab}, N_b معروفة عادة. وفي حالة معاينة عشوائية بسيطة من الإطارين A و B ، اقترح Hartley (1962) التقدير

$$\hat{Y} = \frac{N_A}{n_A} y_a + p\frac{N_A}{n_A} y_{ab} + q\frac{N_B}{n_B} y_{ba} + \frac{N_B}{n_B} y_b \tag{5A.78}$$

حيث $p+q=1$ كما سبق، والمقادير y هي مجاميع العيّنات في الطبقات، ويرمز الدليلان ab و ba للعيّنات في الطبقة المزدوجة الموجودة في n_A و n_B وقد حدد Hartley p و q بحيث يجعل $V(\hat{Y})$ أصغر ما يمكن مع تكلفة مثبتة. وقدّم Lund (1968) و Fuller و (1972) Burmeister تحسينات في تقدير Hartley مستخدمين في الأساس تقديرات لِـ N_a ، N_{ab} ، N_b أفضل مما هو مستخدم في تقدير Hartley وقد عالج Fuller و Burmeister (1972) أيضًا حالة يكون الإطار A فيها مساحيًا، مع معاينة جزئية لوحدات مساحية. ويعطي Hartley (1974) طريقة عامة للمعاينة من إطارين، قابلة للتطبيق في أي تصميم عيّنة من إطارين.

تمارين

(١٥ ـ ١) عند تخطيط مسح إحصائى للمبيعات في نوع معين من المخازن، حيث $n=550$ ، توافرت تقديرات جيدة لِـ S_h من مسح سابق في طبقتين من الطبقات الثلاث. وتتألف الطبقة الثالثة من مخازن جديدة ومخازن لم يكن لها مبيعات في المسح السابق، بحيث يصبح من الضروري تخمين قيمة لِـ S_3 . وإذا كان S_3 مساويًا بالفعل 10 ، فاحسب $V(\bar{y}_{st})$ كما تعطيه محاصّة نيمانية مقدّرة وذلك عندما نخمّن S_3 بِـ (أ) 5 (ب) 20 بيّن أنه في كلتا الحالتين تكون الزيادة المتناسبة في التباين فوق الوضع الأمثل الصحيح أكبر بقليل من 2% .

الطبقة	W_h	S_h الصحيحة	S_h المقدّرة (a)	(b)
1	0.3	30	30	30
2	0.6	20	20	20
3	0.1	10	5	20

(١٥-٢) بيّن أنه إذا كانت جميع المقادير S_h مقدّرة بصورة صحيحة باستثناء S_L المقدّرة على الشكل $\hat{S}_L = S_L(1+\lambda)$, فالزيادة المتناسبة في $V_{opt}(\bar{y}_{st})$, مستخدمين $\hat{S}_L$ بدلاً من القيمة الصحيحة S_L هي في المحاصّة الينمانية .

$$\frac{\lambda^2 n'_L(n-n'_L)}{(1+\lambda)n^2}$$

حيث n'_L حجم العيّنـة في الطبقة L تحت محاصّة نيمانية صحيحة. تحقق من أن هذه العلاقة تتفق مع نتائج التمرين (١٥ ـ ١). (والتوافق ليس مضبوطًا تمامًا بسبب تدوير الـ n_h إلى أعداد صحيحة). وبالتالي بيّن أن التقليل في تقدير S_L بنسبة 50% له نفس أثر مبالغة في التقدير قدرها 100% .

(١٥-٣) في حال وجود طبقتين وإذا كانت ϕ نسبة القيمة الفعلية لِـ n_1/n_2 إلى القيمة النيمانية المثلى لِـ n_1/n_2 فبيّن أنه بصرف النظر عن قيم S_2, S_1, N_2, N_1 لا يمكن أن

تكون النسبة $V_{min}(\bar{y}_{st})/V(\bar{y}_{st})$ أقل من $4\phi(1+\phi)^2$ وذلك عندما يكون الـ ت م قابلاً للإهمال.

(١٥ ـ ٤) يمكن تصنيف نتائج عيّنة عشوائية بسيطة حجمها $n=1000$ إلى ثلاث «طبقات» حيث $\bar{y}_h$ تساوي 10.2 ، 12.6 و17.1 ، و$s_h^2=10.82$ (وهو نفسه في كل طبقة) و $s^2=17.66$. وتقديرات الترجيحات لكل طبقة هي $w_h=0.5, 0.3, 0.2$ ، على الترتيب. ومن المعروف أن هذه الترجيحات ليست مضبوطة تمامًا، إلا أنه يُعتقد أنها جميعًا صحيحة في حدود 5% بحيث تكون أسوأ الحالات إما w_h مساوية لِـ 0.285, 0.525 و 0.190 أو w_h مساوية لِـ 0.315, 0.475 و 0.210 فهل تنصح ، وفقًا لطرق الفقرة (١٥ـ٢)، باستخدام التقسيم إلى طبقات؟ (حيثما تدعو الحاجة ، افترض أن $s_h^2=S_h^2$.) و $\bar{y}_h=\bar{Y}_h$).

(١٥ ـ ٥) الهدف من عيّنة عشوائية طبقية بمتغيرين هو إرضاء الشروط

$$V(\bar{y}_{1st})\leq V_1; \quad V(\bar{y}_{2st})\leq V_2$$

وذلك من أجل تكلفة $C=\sum c_h n_h$. في نهايتها الصغرى. ويمكن تجاهل الـ ت م .

(ا) برهن نتيجة Chatterjee (1972) بأنه لا بد من محاصّة تمثل حلاً وسطًا إذا كان

$$\frac{\sum W_h S_{2h}\sqrt{c_h}}{\sum W_h(S_{1h}^2/S_{2h})\sqrt{c_h}}\leq\frac{V_2}{V_1}\leq\frac{\sum W_h(S_{2h}^2/S_{1h})\sqrt{c_h}}{\sum W_h S_{1h}\sqrt{c_h}}.$$

(ب) إذا كانت V_2/V_1 تساوي أو تتجاوز الحد الأعلى فالمحاصّة المثلى من أجل y_1 تحقق شرطي التساهل كليهما، مع نتائج مقابلة تتعلق بالحد الأدنى.

(١٥ ـ ٦) خططنا مسحًا بثلاث طبقات لتقدير النسبة المئوية للأسر التي تمتلك حسابات في مصارف التوفير ومعدّل المبلغ المستثمر للأسرة الواحدة. وكانت التقديرات المسبقة للنسب المئوية P_h وللتباينات ضمن الطبقات S_h الخاصة بالمبلغ المستثمر كما يلي:

الطبقة	W_h	$P_h(\%)$	$S_h(\$)$
1	0.6	20	90
2	0.3	40	180
3	0.1	70	520

احسب أصغر حجوم للعيّنة n والـ n_h التي تحقق المتطلبات التالية: (ا) تقدير النسبة المئوية للأسر بخطأ معياري $s.e=2$ ، ومتوسط المبلغ المستثمر بخطأ معياري $s.e=\$5$. (ب) تقـدير النسبة المئوية للأسر بخطأ معياري $s.e=1.5$ ، ومتوسط المبلغ المستثمر بخطأ معياري $s.e=\$5$.

الجـزء (ب) يتـطلب محاصّة تمثل حلاً وسطًا، إما باستخدام برمجة حاسب أو بالطريقة المذكورة في الطبقة الثانية، صفحة ١٨٠ . والمحاصّة $n_h=371, 344, 315$ مع $n=1030$ تحقق شرطي التساهل. بيّن أن طريقة Booth-Sedransk (فقرة ١٥ ـ ٤) تعطي $n_h/n=0.431,\ 0.326,\ 0.243$ وستحتاج هذه المحاصّة إلى $n=1073$ لمواجهة شرطي التساهل.

(١٥ ـ ٧) يبيّن الجدول توزيع التكرار لمجتمع من 911 مدينة تقع حجومها بين 10,000 و 60,000 مرتبة وفق فئات طولها 2000 . ولاختصار الحسابات نعطي قيم y' بعد ترميزها وقيم $\sqrt{f}$, . القيم المتجمعة لِـ $\sqrt{f}$, ، القيم المتجمعة لِـ f و fy' و $\Sigma fy'^2$. طبّق قاعـدة Dalenius-Hodjes لإنشاء طبقتين في حالة محاصّة مثلى بالمعنى النيماني. أوجد قيم W_h و S_h لكل من الطبقتين. تحقق من أن (ا) الحجوم المثلى للعيّنة هي نفسها تقريبًا في الطبقتين و(ب) بإيجاد S^2 للمجتمع بكامله تحقق من أن

$$\frac{V(\bar{y})}{V_{opt}(\bar{y}_{st})} \doteq 4.8$$

(١٥ ـ ٨) قسمنا التوزيع المثلثي القائم $f(y)=2(1-y)$ حيث $0<y<1$ إلى طبقتين عند النقطة a .

(ا) بيّن أن

$$W_1 = a(2-a), \qquad W_2 = (1-a)^2$$
$$S_1^{\,2} = \frac{a^2(6-6a+a^2)}{18(2-a)^2}, \qquad S_2^{\,2} = \frac{(1-a)^2}{18}$$

(ب) بيّن أنـه وفقًـا لقـاعـدة القيم المتجمعة لِـ $\sqrt{f}$ يكون أفضل اختيار لِـ a هو $1-1/\sqrt[3]{4}=0.37$ ، وأن القيمة المثلى لِـ n_1/n_2 من أجل هذا الحد بين الطبقتين هي $\frac{27}{25}$

وأن $V(\bar{y}_{st})$ هو حوالي 27% . من القيمة المعطاة في المعاينة العشوائية البسيطة .

(١٥ ـ ٩) في كل من التمرينين (١٥ ـ ٧) و(١٥ ـ ٨)، بيّن أن قاعدة Ekman وهي ثابت $= W_h(y_h - y_{h-1})$ تتفق بفارق بسيط جدًّا مع قاعدة القيم المتجمعة لِـ $\sqrt{f}$ في تعيين حدود الطبقات .

	f	y'	$\sqrt{f}$	القيم المتجمعة لِـ f	القيم المتجمعة لِـ $\sqrt{f}$	fy'
10 –	205	0	14.3	205	14.3	0
12 –	135	1	11.6	340	25.9	135
14 –	106	2	10.3	446	36.2	212
16 –	82	3	9.1	528	45.3	246
18 –	61	4	7.8	589	53.1	244
20 –	42	5	6.5	631	59.6	210
22 –	32	6	5.7	663	65.3	192
24 –	30	7	5.5	693	70.8	210
26 –	27	8	5.2	720	76.0	216
28 –	18	9	4.2	738	80.2	162
30 –	22	10	4.7	760	84.9	220
32 –	21	11	4.6	781	89.5	231
34 –	19	12	4.4	800	93.9	228
36 –	16	13	4.0	816	97.9	208
38 –	14	14	3.7	830	101.6	196
40 –	17	15	4.1	847	105.7	255
42 –	9	16	3.0	856	108.7	144
44 –	8	17	2.8	864	111.5	136
46 –	11	18	3.3	875	114.8	198
48 –	9	19	3.0	884	117.8	171
50 –	7	20	2.6	891	120.4	140
52 –	4	21	2.0	895	122.4	84
54 –	5	22	2.2	900	124.6	110
56 –	5	23	2.2	905	126.8	115
58 –	6	24	2.4	911	129.2	144
Totals	911		129.2			4407

$$\sum fy'^2 = 50{,}395$$

(١٥ـ١٠) يتوافر لدينا مبلغ $5000 من أجل معاينة طبقية، ووفقًا لرموز الفقرة (١٥ـ٨) يُعتقد أن دالة التكلفة هي بصورة تقريبية $C = 200L + 10n$ و

$$V(\bar{y}_{st}) \doteq \frac{S^2}{n}\left[\frac{\rho^2}{L^2} + (1-\rho^2)\right]$$

حيث p الارتباط بين المتغير المستخدم لإنشاء الطبقات والمتغير المراد قياسه في المسح الإحصائي. احسب القيمة المثلى لِـ L حيث p مساوية لِـ 0.95 ، 0.9 و 0.8 . ما هو عدد الطبقات الذي نستخدمه من أجل القيم الثلاث لِـ p والذي يمثل حلًا وسطًا جيدًا؟

(١٥ ـ ١١) البيان التالي مستخلص من عيّنة طبقية من وكلاء الإطارات مأخوذة في آذار (مارس) Deming and Simmons 1945 1946 وقد خُصص الوكلاء إلى طبقات وفقًا لعدد الإطارات الجديدة التي كانت بحوزة كل منهم في تعداد إحصائي سابق. ومتوسطات العيّنة $\bar{y}_h$ هي متوسط عدد الإطارات الجديدة لكل وكيل.

(ا) قدّر الكسب في الدقة العائد إلى التقسيم إلى طبقات.

(ب) قارن هذه النتيجة مع الكسب الذي كنا سنبلغه من محاصّة تناسبية.

حدود الطبقات	N_h	W_h	$\bar{y}_h$	s_h^2	n_h
1–9	19,850	0.8032	4.1	34.8	3000
10–19	3,250	0.1315	13.0	92.2	600
20–29	1,007	0.0407	25.0	174.2	340
30–39	606	0.0245	38.2	320.4	230
المجاميع	24,713	0.9999			4170

(١٥ ـ ١٢) في مجتمع طبقتان حجماهما النسبيان $W_1 = 0.8$, $W_2=0.2$, وتباينات ضمن الطبقات $S_1^2 = 100, S_2^2 = 400$. ونريد أخذ عيّنة عشوائية طبقية تحقق المتطلبات التالية: (i) تقدير متوسط كل طبقة بتباين أصغر أو يساوي الواحد؛ (ii) $V(\bar{y}_{st})\leqslant 0.5$ ومتجاهلين الـ ت م م، أوجد قيم n_1, n_2 التي تحقق المتطلبات الثلاثة جميعًا وذلك من أجل أصغر قيمة ممكنة لِـ $n=n_1+n_2$.

تلميح: لاحظ أن $-\partial V(\bar{y}_{st})/\partial n_1 > -\partial V(\bar{y}_{st})/\partial n_2$ إذا كان $n_1 < 2n_2$. وقد ناقش Fuller (1966) طرقًا مختلفة لمعالجة هذه المسألة.

(١٥ ـ ١٣) في مثال يعود لِـ Nordbotten (1956) وقام بدراسته Kokan (1963) خُطط مسح لتقدير حجم العمالة الكلي Y_1 وقيمة الإنتاج Y_2 في مؤسسات صناعة الأثاث. وعند تقسيم المؤسسات إلى طبقات وفقًا لحجمها، كانت قيم الـ N_h والتقديرات التقريبية لِـ S_{ih}^2 كما يلي:

الطبقة	N_h	S_{1h}^2	S_{2h}^2
كبيرة	600	200	500,000
صغيرة	1,000	10	4,000
	1,600		

ومتطلبات أن لا يكون الخطأ في التقديرين Y_2, Y_1 أكثر من 6% $(P=0.95)$ يؤدي إلى شرطي التساهل

$$V_1 = V(\bar{y}_{1st}) \leq 0.0351: V_2 = V(\bar{y}_{2st}) \leq 56.25$$

بيّن أن المحاصّة المثلى من أجل y_1 حيث $n=617, n_2=167, n_1=450$ تحقق شرطي التساهل كليهما. لاحظ أنه لا يمكن تجاهل الـ ت م م في هذه المسألة.

(١٥ ـ ١٤) في معاينة عشوائية طبقية مع وحدة واحدة في كل طبقة، افترض أنه يكمن تجميع الطبقات في أزواج حيث $N_{j1}=N_{j2}=N_j (j=1,2,\ldots,L/2)$. وفي طريقة بديلة للمعاينة نسحب عشوائيًا وحدتين من كل زوج من الطبقات. بيّن من أجل هذه الطريقة أن:

$$V(\hat{Y}_{st2}) = \frac{(N_j-1)}{(2N_j-1)} \sum_{j=1}^{L/2} \left[2N_j(N_j-1) \sum_{i=1}^{2} S_{ji}^2 + (Y_{j1}-Y_{j2})^2 \right]$$

وبالتالي بيّن أن القيمة المتوقعة لتقدير «الطبقة المنهارة» $v_1(\hat{Y}_{st})$ ، والمذكورة في العلاقة (5A.54) فقرة (١٥ـ١٢)، تبالغ في تقدير $V(\hat{Y}_{st2})$، وهو التباين المناسب في حالة استخدام طبقات حجمها مضاعف.

الفصل السادس

المقدّر النسبة

(٦-١) طرق التقدير

إن أحد مقوّمات نمو الإحصاء النظري هو ظهور قدر كبير من المعالجة النظرية التي تناقش كيفية صنع تقديرات جيدة من بيان إحصائي. وفي تطوير نظرية المعاينة الإحصائية، على وجه التحديد، استُخدم هذا القدر من المعرفة بصورة بسيطة نسبيًا. وأعتقد أن هناك سببين رئيسين لذلك، **الأول** هو أنه في المسوح الإحصائية التي تحوي عددًا كبيرًا من المفردات، توجد فائدة قصوى في طريقة للتقدير لا تتطلب، حتى في الحاسبات، إلا ما يزيد قليلاً على عمليات جمع بسيطة، بينما يمكن أن تضطرنا طرق التقدير المتفوقة في الإحصاء النظري، مثل طريقة الإمكانية العظمى، إلى سلسلة من التقريبات المتتالية قبل أن نتمكن من إيجاد التقدير. **والسبب الآخر**، كما أشرنا في الفقرة (٢-٤)، هو وجود فرق في الموقف بالنسبة لهذين الخطين من البحث العلمي. فمعظم طرق التقدير في الإحصاء النظري تفترض أننا نعرف الشكل الدالّي للتوزيع الاحتمالي الذي تتبعه المعلومات الإحصائية في العيّنة، وتكون طريقة التقدير معدّة بعناية لهذا النوع من التوزيعات. ومن المفضل في نظرية المعاينة أن نضع فروضًا محدودة فقط حول التوزيع الاحتمالي، (مثل أن يكون بعيدًا جدًّا عن التناظر، أو قريبًا إلى حد ما من التناظر)، ونترك التحديد الدقيق للشكل الدالّي خارج إطار المناقشة. ويقود هذا التفضيل إلى استخدام طرق بسيطة تعمل بصورة جيدة تحت عدد من أنواع التوزيعات. وهذا الموقف منطقي فيما يتعلق بمعالجة مسوح إحصائية يتغير فيها التوزيع الاحتمالي من مفردة إلى أخرى، وحيث لا نرغب في التوقف وفحص كل هذه التوزيعات قبل أن نقرر كيفية القيام بكل تقدير.

ونتيجة لذلك، فإن طرق التقدير المتعلقة بأعمال المسح الإحصائي محدودة الآفاق تمامًا في الوقت الراهن. وسنتعرض الآن لطريقتين: طريقة المقدّر النسبة في هذا الفصل وطريقة الانحدار الخطي في الفصل السابع.

(٦-٢) المقدّر النسبة

في طريقة النّسبة، نحصل في كل وحدة من العيّنة على متغير مساعد x_i مرتبط مع y_i. ويجب أن يكون مجموع القياسات x في المجتمع ككل، ولنرمز له بـ X، معروفًا. وفي التطبيق العملي يكون x_i، على الغالب، قيمة y_i في وقت سابق تمّ فيه القيام بحصر شامل. وهدف هذه الطريقة هو الحصول على دقة متزايدة بالاستفادة من الارتباط بين y_i و x_i. وسنفترض الآن أن المعاينة هي معاينة عشوائية بسيطة.

المقدّر النسبة لـ Y مجموع المجتمع y_i هو:

$$\hat{Y}_R = \frac{y}{x}X = \frac{\bar{y}}{\bar{x}}X \tag{6.1}$$

حيث y و x مجموعا العيّنتين من المقادير y_i، x_i على الترتيب.

إذا كانت x_i قيمة y_i في وقت سابق فإن طريقة النسبة تستخدم العيّنة لتقدير التغير النسبي Y/X الذي حدث منذ ذلك الوقت. وبضرب تقدير التغير النسبي، y/x بالمجموع X المعروف في مناسبة سابقة، نحصل على تقدير لمجموع المجتمع في الوقت الراهن. وإذا كانت النسبة y_i/x_i هي تقريبًا نفسها في جميع وحدات المعاينة، فإن قيمة y/x تتغير قليلاً من عيّنة إلى أخرى، ويكون التقدير النسبة ذا دقة عالية. وفي تطبيق آخر، يمكن أن يكون x_i مجموع الفدادين المزروعة في مزرعة و y_i عدد الفدادين المزروعة بمحصول ما. وسيكون التقدير النسبة ناجحًا في هذه الحالة إذا خصص جميع المزارعين حوالي النسبة المئوية نفسها من مجمل زراعاتهم لهذا المحصول.

وإذا كانت الكمية التي نريد تقديرها هي $\bar{Y}$ القيمة المتوسطة للمجتمع y_i فإن التقدير النسبة هو:

$$\hat{Y}_R = \frac{y}{x}\bar{X}$$

وكثيراً ما نرغب في تقدير نسبة* بدلاً من مجموع أو متوسط، مثلاً نسبة فدادين الذرة إلى فدادين القمح، نسبة نفقات العمل إلى مجموع النفقات، أو نسبة الممتلكات المنقولة إلى مجموع الممتلكات. وتقدير العيّنة هو $\hat{R} = y/x$. وفي هذه الحالة لا نحتاج إلى معرفة المجموع X. وقد نوقش استخدام التقديرات النسبة لهذه الغاية في الفقرتين (٢-١١) (المعاينة العنقودية من أجل نسب) و(٣-١٢).

جدول (٦-١) حجوم تسع وأربعين من المدن الكبرى في الولايات المتحدة (بالآلاف) في 1920 (x_i) وفي عام 1930 (y_i)

x_i	y_i	x_i	y_i	x_i	y_i
76	80	2	50	243	291
138	143	507	634	87	105
67	67	179	260	30	111
29	50	121	113	71	79
381	464	50	64	256	288
23	48	44	58	43	61
37	63	77	89	25	57
120	115	64	63	94	85
61	69	64	77	43	50
387	459	56	142	298	317
93	104	40	60	36	46
172	183	40	64	161	232
78	106	38	52	74	93
66	86	136	139	45	53
60	57	116	130	36	54
46	65	46	53	50	58
				48	75

مثـال

يبيّن الجدول (٦-١) عدد السكان (بالآلاف) في كل من عيّنة عشوائية بسيطة من 49 مدينة مسحوبة من مجتمع الـ 196 مدينة المذكورة في الفقرة (٢-١٥). والمسألة هي تقـدير العـدد الكـلي للسكـان في الـ 196 مدينة عام 1930 ونفرض أن المجموع الصحيح X لعام 1920 معروف، وقيمته 22,919.

* تجدر ملاحظة الفرق بين «تقدير نسبة» و«التقدير النسبة». المترجم

وهذا المثال مناسب للتقدير النسبة. وتبيّن معظم المدن في العيّنة زيادة في الحجم بين عامي 1920 و 1930 من مرتبة 20 في المائة. ولدينا من البيان الإحصائي للعيّنة:

وبالتالي فإن التقدير النسبة لمجموع 1930 من أجل الـ 196 مدينة هو

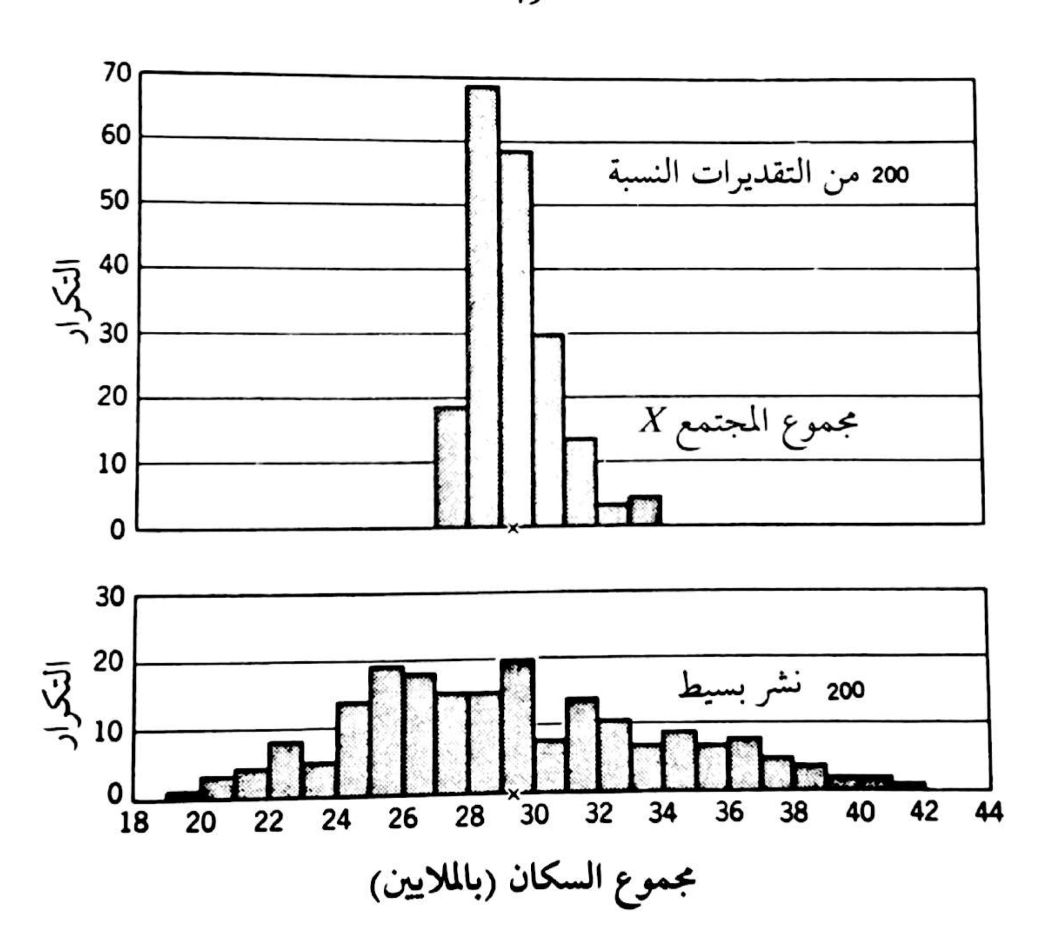
$$y = \sum y_i = 6262, \qquad x = \sum x_i = 5054$$

والتقدير الموافق المبني على متوسط العيّنة للمدينة الواحدة هو

$$\hat{Y}_R = \frac{y}{x} X = \frac{6262}{5054}(22,919) = 28,397$$

والمجموع الصحيح لعام 1930 هو 29,351 .

$$\hat{Y} = N\bar{y} = \frac{(196)(6262)}{49} = 25,048$$



شكل (٦-١) مقارنة تجريبية للتقدير النسبة، حيث يستند التقدير إلى متوسط العيّنة

ويبيّن الشكل (٦-١) التقدير النسبة والتقدير المبني على متوسط العيّنة للمدينة

الـواحـدة، وذلـك لكل من 200 عيّنة عشوائية حجم كل منها 49 مسحوبة من هذا المجتمع. ويتضح وجود تحسّن كبير جدًّا في الدقة من خلال تطبيق طريقة النسبة.

(٦-٣) التباين التقريبي للتقدير النسبة

لقـد أثبت توزيع التقـدير النسبـة أنـه عسـير على المعالجة بشكل مزعج لأن كلًّا من y و x يتغير من عيّنة إلى عيّنة. وتقصرّ النتائج النظرية المعروفة في تقديم كل ما نرغب معرفته في التطبيقات العملية. وسنعرض أولًا النتائج الرئيسة بدون برهان.

والتقدير النسبة متّسق (وهذا واضح). وهو منحاز، باستثناء ما يتعلق بأنواع خاصة من المجتمعات، مع أنه يمكن إهمال الانحياز في العيّنات الكبيرة، وتحت قيود معتـدلة تتعلق بنوع المجتمع الذي نعاينه، ينتهي توزيع التقدير النسبة إلى التوزيع الطبيعي عندما تصبح n كبيرة جدًّا، ويُظهر التوزيع ميلًا إلى عدم التناظر في الاتجاه الموجب في حالة عيّنات معتدلة الحجم، وذلك، على الأقل، في المجتمعات التي غالبًا ما تُستخدم هذه الطريقة من أجلها. ونمتلك قوانين دقيقة تتعلق بالانحياز، ولكننا لا نمتلك لتباين التقدير إلا تقريبات يقتصر جواز تطبيقها على العيّنات الكبيرة.

وتؤدي بنـا هذه النتائج إلى القول بأنه لا توجد صعوبة إذا كانت العيّنة كبيرة بكفـاية بحيث إن (i) توزّع النسبـة هو تقريبًا التوزيع الطبيعي، و (ii) قانون العيّنة الكبـيرة المتعلق بتبـاين النسبـة هو قانـون مشروع. وكقاعدة عمل نقول إنه يمكن استخدام النتائج الموافقة لحالة عيّنة كبيرة إذا كان حجم العيّنة، من جهة، متجاوزًا 30، ومن جهة أخرى كبيرًا بحيث يكون معامل اختلاف كل من $\bar{x}$ و $\bar{y}$ أقل من 10 بالمائة.

نظرية (٦-١)

التقـديرات النسبـة لمجمـوع المجتمع Y ولمتوسط المجتمع $\bar{Y}$ وللنسبة Y/X في المجتمع، هي على الترتيب

$$\hat{Y}_R = \frac{\bar{y}}{\bar{x}}X, \qquad \hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}}\bar{X}, \qquad \hat{R} = \frac{\bar{y}}{\bar{x}}$$

وفي عيّنة عشوائية بسيطة حجمها n (n كبيرة) يكون

$$V(\hat{Y}_R) \doteq \frac{N^2(1-f)}{n}\left[\frac{\sum_{i=1}^{N}(y_i - Rx_i)^2}{N-1}\right] \tag{6.2}$$

$$V(\hat{\bar{Y}}_R) \doteq \frac{1-f}{n}\left[\frac{\sum_{i=1}^{N}(y_i - Rx_i)^2}{N-1}\right] \tag{6.3}$$

$$V(\hat{R}) \doteq \frac{1-f}{n\bar{X}^2}\left[\frac{\sum_{i=1}^{N}(y_i - Rx_i)^2}{N-1}\right] \tag{6.4}$$

حيث $f=n/N$ هو كسر المعاينة. وتبيّن الطريقة المستخدمة في النظرية (٢-٥) أن (6.2), (6.3), (6.4) هي أيضًا تقريبات لمتوسط مربعات الخطأ للكميات المقدّرة المذكورة في هذه العلاقات.

والمناقشة المؤدية إلى النتيجة التقريبية (6.4) كانت قد أُعطيت في النظرية (٢-٥). وبما أن $\hat{Y}_R = N\bar{X}\hat{R}$، $\hat{\bar{Y}}_R = \bar{X}\hat{R}$ فالنتيجتان الباقيتان تتبعان بصورة مباشرة.

نتيجة (١)

توجد أشكال بديلة مختلفة للنتيجة. فيمكن أن نكتب، باعتبار

$$V(\hat{Y}_R = \frac{N^2(1-f)}{n(N-1)}\sum_{i=1}^{N}[(y_i - \bar{Y}) - R(x_i - \bar{X})]^2$$

$$= \frac{N^2(1-f)}{n(N-1)}[\sum(y_i - \bar{Y})^2 + R^2\sum(x_i - \bar{X})^2$$

$$-2R\sum(y_i - \bar{Y})(x_i - \bar{X})]$$

ومعامل الارتباط ρ بين x_i و y_i في مجتمع معرف بالمعادلة

$$\rho=\frac{E(y_i-\bar{Y})(x_i-\bar{X})}{\sqrt{E(y_i-\bar{Y})^2E(x_i-\bar{X})^2}}=\frac{\sum^{N}(y_i-\bar{Y})(x_i-\bar{X})}{(N-1)S_yS_x}$$

وهذا يقود الى النتيجة

$$V(\hat{Y}_R)=\frac{N^2(1-f)}{n}(S_y{}^2+R^2S_x{}^2-2R\rho S_yS_x) \tag{6.5}$$

وهناك شكل مكافىء هو

$$V(\hat{Y}_R)=(1-f)\frac{Y^2}{n}\left(\frac{S_y{}^2}{\bar{Y}^2}+\frac{S_x{}^2}{\bar{X}^2}-\frac{2S_{yx}}{\bar{Y}\bar{X}}\right) \tag{6.6}$$

حيث $S_{yx}=\rho S_yS_x$ هو التغاير بين y_i و x_i. ويمكن كتابة هذه العلاقة أيضًا على الشكل

$$V(\hat{Y}_R)=(1-f)\frac{Y^2}{n}(C_{yy}+C_{xx}-2C_{yx}) \tag{6.7}$$

حيث C_{yy} ، C_{xx} هما مربعا معامل اختلاف y_i و x_i على الترتيب، و C_{yx} هو التغاير النسبي .

نتيجة (٢)

بما أن $\hat{Y}_R$ ، $\hat{\bar{Y}}_R$ و $\hat{R}$ لا تختلف إلا بمعاملات معروفة، فيكون معامل الاختلاف (أي الخطأ المعياري مقسومًا على الكمية المقدّرة) نفســه من أجل التقديرات الثلاثة . ومن (6.7) نجد أن مربع معامل الاختلاف هذا هو

$$(\text{cv})^2=\frac{V(\hat{Y}_R)}{Y^2}=\frac{1-f}{n}(C_{yy}+C_{xx}-2C_{yx}) \tag{6.8}$$

ويسمّي Hansen وآخرون (1953) مربع معامل الاختلاف $(cv)^2$ التباين النسبي . ويجنّبنا استخدامه تكرار علاقات التباين من أجل مقادير على صلة ببعضها مثل تقدير مجموع المجتمع وتقدير متوسطه .

(٦-٤) تقدير التباين من عيّنة

من المعادلة (6.2) نجد

$$V(\hat{Y}_R) \doteq \frac{N^2(1-f)}{n} \frac{\sum^N (y_i - Rx_i)^2}{N-1}$$

وكما ذكرنا سابقًا في الفقرة (٢-١١) نأخذ

$$\frac{\sum^n (y_i - \hat{R}x_i)^2}{(n-1)}$$

كتقدير عيّنة لتباين المجتمع. ولهذا التقدير انحياز من مرتبة $1/n$.
ويعطي هذا في حالة تقدير التباين، $v(\hat{Y}_R)$:

$$v(\hat{Y}_R) = \frac{N^2(1-f)}{n(n-1)} \sum_{i=1}^{n} (y_i - \hat{R}x_i)^2 \tag{6.9}$$

ويمكن التعبير عن هذه النتيجة بعدة طرق مختلفة. وعلى سبيل المثال،

$$v(\hat{Y}_R) = \frac{N^2(1-f)}{n(n-1)} \left(\sum y_i^2 + \hat{R}^2 \sum x_i^2 - 2\hat{R} \sum y_i x_i\right) \tag{6.10}$$

$$= \frac{N^2(1-f)}{n} \left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R} s_{yx}\right) \tag{6.11}$$

حيث $s_{yx} = \sum (y_i - \bar{y})(x_i - \bar{x})/(n-l)$ هو تغاير العيّنة بين y_i و x_i.

وتوجد علاقتان بديلتان من أجل تقدير التباين من عيّنة. وبما أن $\hat{y}_R = N\bar{X}\hat{R}$ ، فأحد الأشكال من أجل $\hat{R}$ هو

$$v_1(\hat{R}) = \frac{(1-f)}{n\bar{X}^2} \left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R} s_{yx}\right) \tag{6.12}$$

وعلى أي حال، وباعتبار $\hat{R} = \bar{y}/\bar{x}$، فلا حاجة لمعرفة الكمية $\bar{X}$ وأحيانًا لا تكون معروفة عند تقدير R . وهذا يقترح الشكل البديل،

$$v_2(\hat{R}) = \frac{(1-f)}{n\bar{x}^2}(s_y{}^2 + \hat{R}^2 s_x{}^2 - 2\hat{R}s_{yx}) \tag{6.13}$$

ويمكن استخدام هذا الشكل أيضًا من أجل $v(\hat{Y}_R)$ آخذين $v_2(\hat{Y}_R) = X^2 v_2(\hat{R})$.

وهذا يطرح السؤال التالي: إذا كان $\bar{X}$ معروفًا فهل v_1 أفضل من v_2؟ والجواب غير واضح حاليًا. وقد درس P.S.R.S. Rao و J.N.K. Rao (1971) بصورة تحليلية الانحياز في كل من v_1 و v_2 وذلك في مجتمعات منتهية، مفترضين نموذج الانحدار الخطي،

$$y_i = \alpha + \beta x_i + e_i$$

مع

$$E(e_i|x_i) = 0,\ V(e_i|x_i) = \delta x_i{}^t,\ E(e_i e_j|x_i x_j) = 0,$$

حيث $2 \leq t \leq 0$، وتوزيع x_i هو توزيع جاما $ax^{h-1}e^{-x}$. وقد دُرس المدى $0 \leq t \leq 2$ لأنه يُعتقد في التطبيقات أن راسب التباين لِـ y_i يزداد مع x_i وفق معدلات تختلف باختلاف المجتمع. وقد وجدا أن v_2 أقل انحيازًا من أجل $0 \leq t \leq 3/2$ إلا أنه أقل استقرارًا أيضًا من أجل $t=0$ أو $t=1$.

(٦-٥) حدود ثقة

إذا كان حجم العيّنة كبيرًا إلى الحد الذي يكفي لتطبيق التقريب الطبيعي فيمكن الحصول على حدود ثقة من أجل Y و R كما يلي:

$$Y:\ \hat{Y}_R \pm z\sqrt{v(\hat{Y}_R)} \tag{6.14}$$

$$R:\ \hat{R} \pm z\sqrt{v(\hat{R})} \tag{6.15}$$

حيث z قيمة المتغير الطبيعي المعياري الموافقة لاحتمال الثقة الذي اخترناه.

وقد اقتُرح في الفقرة (٦-٣) أن تطبيق التقريب الطبيعي يكون مقبولاً إذا كان حجم العيّنة لا يقلّ عن 30 ، وكان من الكبر بحيث يجعل معامل اختلاف $\bar{y}$ و $\bar{x}$ أقل

من 0.1 . وعندما لا تنطبق هذه الشروط تميل العلاقة الخاصة بـ $v(\hat{R})$ لإنتاج قيم منخفضة جدًّا، كما قد يصبح الالتواء الموجب في توزيع $\hat{R}$ ملحوظًا.

وقـد استُخدمت طرق بديلة لحساب حدود الثقة تأخذ بعض الاعتبار لالتواء توزيع $\hat{R}$ وذلـك في بحث بيولـوجي Fieller ,(1932) و Paulson (1942) . وتتـطلب الطريقة أن يتبع $\bar{y}$ و $\bar{x}$ التوزيع الطبيعي الثنائي بحيث يتوزع $(\bar{y}-R\bar{x})$ وفق التوزيع الطبيعي. وينتج من ذلك أن توزيع الكمية

$$\frac{\bar{y}-R\bar{x}}{\sqrt{[(N-n)/Nn]}\sqrt{s_y{}^2+R^2s_x{}^2-2Rs_{yx}}} \tag{6.16}$$

هو تقريبًا التوزيع الطبيعي بمتوسط يساوي الصفر وانحراف معياري يساوي الواحد.

وقيمة R مجهولة، إلا أنه يمكن اعتبار أي قيمة تأملية لِـ R ، تجعل هذا المتغير الطبيعي كبيرًا بكفاية، كقيمة يرفضها البيان الإحصائي للعيّنة. وبالتالي فإنه يمكن إيجـاد حدود الثقة لِـ R بوضع (6.16) مساوية $\pm z$ ، وحل المعادلة الناتجة من الدرجة الثانية في R . وحدود الثقة الناتجة تقريبية لأن جذري المعادلة من الدرجة الثانية تخيليان من أجـل بعض العيّنات. وتصبح مثل هذه الحالات نادرة إذا كان معامل اختلاف $\bar{y}$ و $\bar{x}$ أقل من 0.3 .

ويمكن بعد التبسيط التعبير عن الجذرين على الشكل:

$$R=\hat{R}\frac{(1-z^2c_{\bar{y}\bar{x}})\pm z\sqrt{(c_{\bar{y}\bar{y}}+c_{\bar{x}\bar{x}}-2c_{\bar{y}\bar{x}})-z^2(c_{\bar{y}\bar{y}}c_{\bar{x}\bar{x}}-c_{\bar{y}\bar{x}}{}^2)}}{1-z^2c_{\bar{x}\bar{x}}} \tag{6.17}$$

حيث

$$c_{\bar{y}\bar{y}}=\frac{N-n}{Nn}\frac{s_y{}^2}{\bar{y}^2}$$

هو مربع تقدير معامل اختلاف $\bar{y}$ مع تعاريف مشابهة لِـ $c_{\bar{y}\bar{x}}$ و $c_{\bar{x}\bar{x}}$. وإذا كانت المقادير $z^2c_{\bar{y}\bar{y}}$، $z^2c_{\bar{x}\bar{x}}$، و $z^2c_{\bar{y}\bar{x}}$ ، كلها صغيرة بالنسبة للواحد، فيمكن اختصار عبارة حدود الثقة لتصبح:

$$R = \hat{R} \pm z\sqrt{c_{\bar{y}\bar{y}} + c_{\bar{x}\bar{x}} - zc_{\bar{y}\bar{x}}}$$

وهذه العبارة هي عبارة التقريب الطبيعي في (6.15) نفسها.

وحتى في حالة توزيع طبيعي ثنائي انتُقدت حدود Fieller بأنها ليست محافظة بشكل كاف. ويشرح James, Wilkinson و Venables (1975) طبيعة الصعوبة ويقدمون طريقة بديلة.

(٦-٦) مقارنة التقدير النسبة بالمتوسط لكل وحدة

درسنا في الفصول السابقة نوعًا من التقدير لـ Y هو $N\bar{y}$، حيث $\bar{y}$ متوسط العيّنة (في معاينة عشوائية بسيطة) أو المتوسط المرجّح (في معاينة عشوائية طبقية). وستدعى التقديرات من هذا النوع بالتقديرات المبنيّة على المتوسط لكل وحدة أو التقديرات الناتجة عن النشر البسيط.

نظرية (٦-٢)

في العيّنات الكبيرة ومع معاينة عشوائية بسيطة، يكون تباين التقدير النسبة $\hat{Y}_R$ أصغر من تباين التقدير $\hat{Y} = N\bar{y}$ الذي نحصل عليه بالنشر البسيط، إذا كان:

$$\rho > \frac{1}{2}\left(\frac{S_x}{\bar{X}}\right) \Big/ \left(\frac{S_y}{\bar{Y}}\right) = \frac{\text{معامل اختلاف } x_i}{\text{معامل اختلاف } y_i}$$

برهان

لدينا من أجل $\hat{Y}$

$$V(\hat{Y}) = \frac{N^2(1-f)}{n} S_y{}^2$$

ومن أجل التقدير النسبة لدينا من (6.5)

$$V(\hat{Y}_R) = \frac{N^2(1-f)}{n}(S_y{}^2 + R^2 S_x{}^2 - 2R\rho S_y S_x)$$

ومنه فإن تباين التقدير النسبة سيكون أصغر إذا كان

$$S_y^2 + R^2 S_x^2 - 2R\rho S_y S_x < S_y^2$$

وإذا كان $R = \bar{Y}/\bar{X}$ موجبًا يصبح هذا الشرط

$$\rho > \frac{RS_x}{2S_y} = \frac{1}{2}\left(\frac{S_x}{\bar{X}}\right) \bigg/ \left(\frac{S_y}{\bar{Y}}\right) = \frac{1}{2}\,\frac{\mathrm{cv}(x)}{\mathrm{cv}(y)} \qquad (6.18)$$

(٦-٧) الشروط التي يكون المقدر النسبة تحتها أفضل مقدر خطي غير منحاز

تشير إحدى النتائج المعروفة جيدًا في نظرية الانحدار إلى نوع المجتمع الذي يكون التقدير النسبة من أجله هو التقدير الأفضل من بين صف واسع من التقديرات. وقد بُرهنت النظرية أولاً من أجل المجتمعات اللانهائية ثم قام Brewer (1963 b) و Royall (1970 a) بتمديد هذه النتائج إلى حالة المجتمعات النهائية وتصح هذه النتيجة إذا توافر شرطان.

١ ـ العلاقة بين y_i و x_i هي علاقة خط مستقيم يمر من المبدأ.

٢ ـ تباين y_i حول هذا الخط متناسب مع x_i.

ونعرّف «أفضل تقدير خطي غير منحاز» كما يلي: لنعتبر جميع التقديرات $\hat{Y}$ لـ Y التي هي دوال خطية في قيم العيّنة y_i، أي من الشكل

$$l_1 y_1 + l_2 y_2 + \cdots + l_n y_n$$

حيث لا تعتمد المقادير l_i على y_i مع أنها يمكن أن تكون دوال في x_i، ولنقتصر عند اختيار المقادير l_i على القيم التي تؤدي إلى تقديرات غير منحازة لـ Y . فمن بين هذه التقديرات الخطية غير المنحازة يكون التقدير ذو التباين الأصغر هو «أفضل تقدير خطي غير منحاز» (ا ت خ غ).

ويفترض Brewer و Royall أن جميع قيم المجتمع (y_i, x_i) وعددها N هي عيّنة

عشوائية من مجتمع فوقي يكون فيه

$$y_i = \beta x_i + \varepsilon_i \tag{6.19}$$

حيث الـ ε_i مستقلة عن الـ x_i و $x_i>0$. وفي البيانات التي يكون فيها x_i مثبتًا، نفترض لـ x_i متوسطًا يساوي الصفر وتباينًا يساوي λx_i . والمقادير $x_i(i=1,2,\ldots,N)$ معروفة.

وفي النظرية العشوائية المستخدمة حتى الآن في هذا الكتاب، اعتُبر مجموع مجتمع منته Y كمية مثبتة. وتحت النموذج (6.19) تكون $Y=\beta X+\sum^{N}\varepsilon_i$ على الوجه الآخر، متغيرًا عشوائيًا. وعند تعريف مقدِّر غير منحاز تحت هذا النموذج يستخدم Brewer و Royall مفهومًا لعدم الانحياز يختلف عن ذلك الذي نجده في النظرية العشوائية. ويعتبر أن المقدر $\hat{Y}$ مقدرًا غير منحاز إذا كان $E(\hat{Y})$ مساويًا لـ $E(Y)$ في اختيارات مكررة للمجتمع المنتهي وللعيّنة تحت النموذج المفروض. ويمكن تسمية مقدِّر كهذا نموذج ـ لا انحياز.

نظرية (٦-٣)

تحت النموذج (6.19) يكون المقدر النسبة $\hat{Y}_R = X\bar{y}/\bar{x}$ أفضل مقدر خطي غير منحاز من أجل أي عيّنة، عشوائية كانت أم لا، نختارها وفقًا لقيم المقادير x_i دون غيرها.

برهان

بما أن $E(\varepsilon_i|x_i)=0$ عند تكرار المعاينة، فنستنتج من (6.19) أن

$$Y=\beta X+\sum^{N}\varepsilon_i : E(Y)=\beta X \tag{6.20}$$

وفضلاً عن ذلك، يكون أي مقدر خطي $\hat{Y}$ تحت النموذج (6.19) من الشكل،

$$\hat{Y}=\sum^{n} l_i y_i = \beta \sum^{n} l_i x_i + \sum^{n} l_i \varepsilon_i \tag{6.21}$$

وإذا احتفظنا بقيم العيّنة x_i الـ n مثبتة عند تكرار المعاينة تحت النموذج (6.19) نجد،

$$E(\hat{Y}) = \beta \sum^{n} l_i x_i : V(\hat{Y}) = \lambda \sum^{n} l_i^2 x_i \tag{6.22}$$

ومن (6.20) و (6.22) نجد بوضوح أن $\hat{Y}$ نموذج ـ لا انحياز إذا كان $\sum^{n} l_i x_i = X$. ويجعل $V(\hat{Y})$ أصغر ما يمكن تحت هذا الشرط، مستخدمين طريقة مضاريب لاغرانج، نجد

$$2 l_i x_i = c x_i : l_i = \text{ثابت} = X/n\bar{x} \tag{6.23}$$

ويجب أن تكون قيمة الثابت مساوية لِـ $X/n\bar{x}$ كي يتحقق شرط النموذج ـ لاانحياز وهو $l \sum^{n} x_i = X$. وبالتالي يكون المقدر الـ (ا ت خ غ) $\hat{Y}$ مساويًا: لِـ $\frac{n\bar{y}X}{n\bar{x}} = X\bar{y}/\bar{x} = \hat{Y}_R$, وهو المقدِّر النسبة المعتاد. وهذا يكمل البرهان.

وفضلاً عن ذلك، نجد من (6.20) و (6.21) حيث $l = X/n\bar{x}$

$$\hat{Y}_R - Y = \sum^{n} l_i \varepsilon_i - \sum^{N} \varepsilon_i = (X/n\bar{x})(\sum^{n} \varepsilon_i) - \sum^{N} \varepsilon_i \tag{6.24}$$

$$= \frac{(X - n\bar{x})}{n\bar{x}} \sum^{n} \varepsilon_i - \sum^{N-n} \varepsilon_i \tag{6.25}$$

وحيث يرمز $\sum^{N-n}$ للمجموع فوق قيم المجتمع الـ $(N\text{-}n)$ غير الموجودة في العيّنة. وبالتالي،

$$V(\hat{Y}_R) = \frac{\lambda (X - n\bar{x})^2 (n\bar{x})}{(n\bar{x})^2} + \lambda (X - n\bar{x}) = \frac{\lambda (X - n\bar{x}) X}{n\bar{x}} \tag{6.26}$$

ويمكن بسهولة البرهان على أن تقدير نموذج ـ لاانحياز لِـ λ من هذه العيّنة هو

$$\hat{\lambda} = \sum^{n} \frac{1}{x_i} (y_i - \hat{R} x_i)^2 / (n-1) \tag{6.27}$$

حيث $\hat{R} = \bar{y}/\bar{x}$, كالمعتاد. ويمكن تعويض هذه القيمة في (6.26) لتعطي تقدير عيّنة نموذج ـ لاانحياز لِـ $V(\hat{Y}_R)$.

والميزة العملية لهذه النتيجة هي أنها تقترح الشروط التي لا يكون التقدير النسبة تحتها متفوقًا على المتوسط لكل وحدة فقط، ولكنه التقدير الأفضل من بين صف كامل من التقديرات. وعندما نحاول تقرير نوع التقدير الذي سنستخدمه، فسيكون من المفيد رسم خط بياني لقيم العينة y_i في مقابل x_i. وإذا أظهر هذا الرسم البياني علاقة خط مستقيم يمر من المبدأ، وبدا تباين النقاط y_i حول الخط وكأنه يزداد بصورة متناسبة تقريبًا مع x_i فسيكون من الصعب قهر التقدير النسبة.

وأحيانًا لا يكون تباين y_i في بيانات ثبّتنا فيها x_i متناسبًا مع x_i. وإذا كان راسب التباين من الشكل $\lambda v(x_i)$ حيث $v(x_i)$ معروف، فقد بيّن Brewer و Royall أن المقدر الـ(ا ت خ غ) يصبح

$$\hat{Y} = X\frac{\sum^{n} w_i y_i x_i}{\sum^{n} w_i x_i^2} \tag{6.28}$$

حيث $w_i = 1/v(x_i)$. وقد وجد Jessen وآخرون (1947) في عيّنة سكانية من اليونان أن راسب التبـاين يتـزايد تقـريبًا كتزايد x_i^2 . ويقترح هذا انحدارًا مرجّحًا يكون فيه $w_i = 1/x_i^2$ مما يعطي

$$\hat{Y} = \frac{X\left(\sum^{n} w_i y_i x_i\right)}{\sum^{n} (w_i x_i^2)} = \frac{X}{n}\sum^{n}\left(\frac{y_i}{x_i}\right) \tag{6.29}$$

ومن أجـل مجتمع معطى و n معطاة، وبافتراض أن كل $x_i>0$ يتضح أن قيمة $V(\hat{Y}_R)$ المعـطاة في (6.26) ستكـون أصغـر ما يمكن عنـدما تتألف العيّنة من الـ n مقدارًا x_i الأكبر في المجتمع. وفي 16 من المجتمعات الطبيعية الصغيرة من النوع الذي تُطبق فيه التقـديرات النسبـة وجـد Royall (1970)، من أجـل عيّنـات حجمها n يساوي 2 إلى 12، أن اختيار الـ n مقدارًا x_i الأكبر قد زاد عادة من دقة $\hat{Y}_R$

والخلاصة، تبيّن نتائج Brewer-Royall أن افتراض نوع معين من النماذج يقود

إلى التقـدير النسبـة غير المنحـاز وإلى علاقـات لِـ$V(\hat{Y}_R)$ و $v(\hat{Y}_R)$ بسيطة ومضبـوطة في حالة n أكبر من الواحد. ويمكن استخدام النتائج عمليًا لحالات نجد فيها، لدى تفحّص الأزواج y,x ، في البيان الإحصـائي المتـوافر، أن النموذج صحيح إلى حد مقبول. وتبدو علاقات التباين (6.26) و (6.27) حساسة لأي عدم دقة في النموذج، مع أن هذه المسألة تحتاج إلى مزيد من الدراسة.

ويناقش عمـل آخـر لِـ Royall و Herson (1973) نوع توزيع المعاينة الذي نحتاجه بالنسبـة للمتغيرات x_i كي يبقى $\hat{Y}_R$ غير منحاز وذلك في حالة انحدار كثيرة حدود لِـ y_i على x_i .

(٦ ـ ٨) انحياز التقدير النسبة

للتقـدير النسبـة بصورة عامة، انحياز من مرتبة $1/n$. وبما أن الخطأ المعياري للتقدير من مرتبة $\frac{1}{\sqrt{n}}$. فتكون الكمية $\left(\frac{\text{الانحياز}}{\text{الخطأ المعياري}}\right)$ من مرتبة $\frac{1}{\sqrt{n}}$ أيضًا وتكون مهملة عندما يصبح n كبيرًا. وعادة، لا تكون هذه الكمية مهمة، عمليًا، في عيّنات معتدلة الحجم. إلا أن قيمتهـا، في عيّنـات صغـيرة في معـاينـة طبقية تتضمن العديد من الطبقات، لا تخلو من الأهمية، إذ قد نرغب في حساب ودراسة التقديرات النسبة في طبقات بمفردها حيث العيّنات المأخوذة من هذه الطبقات صغيرة. ونقدّم هنا نتيجتين هامتين حول الانحياز.

وتعطي الأولى الحد الرئيس في عبارة الانحياز عند فكّها وفق سلسلة تايلور.

$$\hat{R}-R=\frac{\bar{y}}{\bar{x}}-R=\frac{\bar{y}-R\bar{x}}{\bar{x}}$$

لنكتب:

$$\frac{1}{\bar{x}}=\frac{1}{\bar{X}+(\bar{x}-\bar{X})}=\frac{1}{\bar{X}}\left(1+\frac{\bar{x}-\bar{X}}{\bar{X}}\right)^{-1}\doteq\frac{1}{\bar{X}}\left(1-\frac{\bar{x}-\bar{X}}{\bar{X}}\right) \qquad (6.30)$$

وبالتالي،

$$\hat{R}-R\doteq\frac{\bar{y}-R\bar{x}}{\bar{X}}\left(1-\frac{\bar{x}-\bar{X}}{\bar{X}}\right) \qquad (6.31)$$

والآن،

$$E(\bar{y}-R\bar{x}) = \bar{Y} - R\bar{X} = 0$$

وهكذا يأتي الحد الرئيس للانحياز من الحد الثاني داخل القوسين. وفضلاً عن ذلك،

$$E\bar{y}(\bar{x}-\bar{X}) = E(\bar{y}-\bar{Y})(\bar{x}-\bar{X}) = \frac{1-f}{n}\rho S_y S_x \qquad (6.32)$$

وذلك استنادًا إلى النظرية (٢-٣) (صفحة ٣٧) وتعريف ρ. وأيضًا،

$$E\bar{x}(\bar{x}-\bar{X}) = E(\bar{x}-\bar{X})^2 = \frac{1-f}{n}S_x^2$$

وبالتالي يكون الحد الرئيس للانحياز

$$E(\hat{R}-R) \doteq \frac{1-f}{n\bar{X}^2}(RS_x^2 - \rho S_y S_x) \qquad (6.33)$$

$$\doteq \frac{1-f}{n}(C_{xx} - C_{yx})R \qquad (6.34)$$

ومن أجل تبرير دقيق لـ(6.34) انظر David وSukhatme (1974). والآن يكون الحد الرئيس في $V(\hat{R})$ هو

$$V(\hat{R}) = \frac{(1-f)}{n\bar{X}^2}(S_y^2 + R^2 S_x^2 - 2R\rho S_y S_x) \qquad (6.35)$$

وذلك من (6.5) بعد تعويض $\hat{R} = \hat{Y}_R / N\bar{X}$.

ومن (6.33) و(6.35) يمكن التعبير عن الحد الرئيس للكمية $\left(\frac{\text{الانحياز}}{\text{الخطأ المعياري}}\right)$، وهي نفسها من أجل $\hat{R}$، $\hat{Y}_R$، و $\hat{\bar{Y}}_R$ على الشكل،

$$\frac{\text{الانحياز}}{\text{الخطأ المعياري}} = \text{cv}(\bar{x})\frac{(RS_x - \rho S_y)}{(R^2 S_x^2 - 2R\rho S_y S_x + S_y^2)^{1/2}} \qquad (6.36)$$

حيث $\text{cv}(\bar{x}) = \sqrt{1-f}S_x/\sqrt{n}\bar{X}$. وبتعويض تقديرات العيّنة للحدود في (6.36) قام Kish، Namboodiri وPillian (1962) بحساب قيم $\left(\frac{\text{الانحياز}}{\text{الخطأ المعياري}}\right)$ لعدد كبير من المفردات في دراسات قومية ومحلية مختلفة. وفي الدراسات القومية كانت جميع قيم

$\left(\frac{\text{الانحياز}}{\text{الخطأ المعياري}}\right)$ تقريبًا أصغر من 0.03 ، وعلى وجه التقريب كانت القيم الوحيدة الأكبر من 0.10 في دراساتهم من أجل طبقة بمفردها تتضمن $n_h=6$ مستشفيات صغيرة.

والنتيجة الثانية التي تعود إلى Hartley وRoss (1954) تعطي نتيجة مضبوطة من أجـل الانحياز، وحـدًا أعـلى لنسبة الانحياز إلى الخطأ المعياري. لنعتبر الآن تغاير $\hat{R}$ و $\bar{x}$ في عيّنات عشوائية بسيطة حجمها n . فلدينا،

$$\operatorname{cov}(\hat{R}, \bar{x}) = E\left(\frac{\bar{y}}{\bar{x}}\cdot\bar{x}\right) - E(\hat{R})\,E(\bar{x}) \tag{6.37}$$

$$= \bar{Y} - \bar{X}E(\hat{R}) \tag{6.38}$$

ومنه

$$E(\hat{R}) = \frac{\bar{Y}}{\bar{X}} - \frac{1}{\bar{X}}\operatorname{cov}(\hat{R}, \bar{x}) = R - \frac{1}{\bar{X}}\operatorname{cov}(\hat{R}, \bar{x}) \tag{6.39}$$

وهكـذا يكون الانحياز في $\hat{R}$ مساويًا $-\operatorname{cov}(\hat{R}, \bar{x})/\bar{X}$. وهذه العبارة مضبوطة خلافًا لتقريب تايلور في (6.33) من أجل الانحياز.

وفضلًا عن ذلك

$$|\text{الانحياز في } \hat{R}| = \frac{|\rho_{\hat{R},\bar{x}}\sigma_{\hat{R}}\sigma_{\bar{x}}|}{\bar{X}}$$

$$\leq \frac{\sigma_{\hat{R}}\sigma_{\bar{x}}}{\bar{X}}$$

باعتبار أنه لا يمكن أن يكون لـ $\hat{R}$ و $\bar{x}$ ارتباط أكبر من الواحد. وبالتالي،

$$\frac{|\text{الانحياز في } \hat{R}|}{\sigma_{\hat{R}}} \leq \frac{\sigma_{\bar{x}}}{\bar{X}} = \text{معامل اختلاف } \bar{x} \tag{6.40}$$

وبالطبع ينطبق الحد الأعلى نفسه على الانحياز في $\hat{Y}_R$ و $\hat{\bar{Y}}_R$. وهكذا إذا كان معامل اختلاف $\bar{x}$ أقل من 0.1 ، فيمكن الاطمئنان إلى اعتبار الانحياز مهملًا بالنسبة للخطأ المعياري.

(٦ - ٩) دقــة العــلاقــات
الخاصة بالتباين وتقدير التباين

وفي حالة عيّنات صغيرة، مثلاً $n<30$ وC_{xx} كبير، كان يُشك منذ زمن طويل في أن العــلاقــات الخــاصــة بالعيّنــات الكبــيرة لكل من $V(\hat{R})$ و$v(\hat{R})$ ستؤدي إلى تقدير بالنقصان. ومستخدمًا النشر وفق سلسلة تايلور، عبّر Sukhatme (1954) عن الخطأ في $V(\hat{R})$ بدلالة العزوم الثنائية لـy وx. ومن سوء الحظ فإن النتيجة هي من التعقيد بحيث لا تقود إلى دليل مفيد في التطبيقات العملية.

وإذا كان y وx يتبعان التوزيع الطبيعي الثنائي فيمكن تبسيط نتيجة Sukhatme بصورة كبيرة. ليكن

$$V_1 = (C_{yy} + C_{xx} - 2C_{yx})/n$$

رمزًا للتقريب الأول للتباين النسبي لـ$\hat{R}$، متجاهلين الـ ت م م. فلحدود من مرتبة $\frac{1}{n^2}$ نجد،

$$E\left(\frac{\hat{R}-R}{R}\right)^2 \doteq V_1\left(1+\frac{3C_{xx}}{n}+\frac{6C_{xx}}{n}.\frac{\rho^2 C_{yy}+C_{xx}-2C_{yx}}{C_{yy}+C_{xx}-2C_{yx}}\right) \tag{6.41}$$

وبما أن الحد الأيمن ضمن القوسين أقل من $6C_{xx}/n$ فهذا يعطي

$$E\left(\frac{\hat{R}-R}{R}\right)^2 < V_1\left(1+\frac{9C_{xx}}{n}\right) \tag{6.42}$$

وذلــك إلى حدود من مرتبــة $\frac{1}{n^2}$. والآن C_{xx}/n هو مربــع معــامــل اختــلاف $\bar{x}$. وبــالتــالي إذا كان n كبيراً بكفاية بحيث يكون معامل اختلاف $\bar{x}$ أقل من 0.1، فإن استخــدام V_1 ينبغي أن لا يؤدي إلى نقص في تقــدير القيمــة الصحيحــة يتجــاوز الـ9% وعمليًا، يبــدو المعــامــل 9 في (6.42) مرتفعًا أكثر من الــلازم بالمقــارنــة مع (6.41). وعلى سبيل المثال، إذا كان $C_{xx}=C_{yy}$ تُختزل (6.41) إلى

$$E\left(\frac{\hat{R}-R}{R}\right)^2 \doteq V_1\left[1+\frac{C_{xx}}{n}(6-3\rho)\right] \tag{6.43}$$

وبـما أن ρ موجب دائـمًا تقـريبًا في تطبيقات طريقة النسبة فإن معاملاً بين 3 و 6 سيكون معقولاً أكثر. وعلى أي حال فإن تأثيرات عدم كون x و y طبيعيين يدخل أيضًا ضمن الحد من مرتبة $\frac{1}{n^2}$.

ومن دراسة وفقًا لطريقة مونت كارلو أجراها Rao (1968) على مجتمعات واقعية صغيرة، سنقتبس بعض النتائج التوضيحية حول انحيازات علاقات العيّنات الكبيرة الخـاصـة بـ $V(\hat{R})$ و $v_1(\hat{R})$ وذلك في حالة عيّنات عشوائية بسيطة من ثمانية مجتمعات فيهـا $N>30$. والمجتمعـات موصـوفـة في Rao (1969) والعـلاقـات لكـل من $V(\hat{R})$ و $v_1(\hat{R})$ المعـطاة في (6.4) و (6.12) ستقـوّم كمقـدّرات للقيمـة الصحيحـة لـ $MSE(\hat{R})$

ومن أجل مجتمع معطى تكون الكميتان $100[MSE(\hat{R}) - V(\hat{R})]/MSE(\hat{R})$ و $100[MSE(\hat{R}) - Ev_1(\hat{R})]/MSE(\hat{R})$ النسبتين المئويتين للنقص في تقدير القيمة الصحيحة لـ $MSE(\hat{R})$. ومتوسطات هذه النسب المئوية في المجتمعات الثمانية مبيّنة في الجدول (٦-٢).

جدول (٦ ـ ٢) متوسط النسبة المئوية للنقص في تقدير $MSE(\hat{R})$

المقدّر	n = 4	6	8	12
$V(\hat{R})$ في (6.4)	14	14	14	12
$v_1(\hat{R})$ في (6.12)	31	23	21	18

وفي هذه البيانات، من النادر إطلاقًا أن تهبط النسبة المئوية للنقص في تقدير $V(\hat{R})$ مع زيادة n . وهـذا مفسّر جزئيًا بواقعـة أنـه في أحـد المجتمعـات أعطى $V(\hat{R})$ زيادة في التقـدير انخفضت قيمتهـا مع انخفـاض n . وتشـير هذه النتـائج إلى أن الانحيازات في $v_1(\hat{R})$ هي، في العيّنـات الصغـيرة، أكثـر جدية بكثـير مما هي في $\hat{R}$ نفسهـا، وهي غير مُرضية حتى القيمـة $n=12$ على الأقـل. وعنـدمـا $n=4$ وجـد Koop (1968) نقصًا في تقديرات $v_1(\hat{R})$ متوسطه 25% في ثلاثة مجتمعات حجم كل منها $N=20$. والمقدّر البديل $v_1(\hat{R})$ الذي يدعو للتفاؤل بالنسبة لتخفيض الانحياز معروض في الفقرة (٦-١٧).

(٦-١٠) التقديرات النسبة
في معاينة عشوائية طبقية

توجد طريقتان يمكننا بوساطتهما الحصول على التقدير النسبة لمجموع المجتمع Y. إحداهما هي أن نحسب التقديرات النسبة منفصلة لمجموع كل طبقة ثم نجمع هذه المجاميع. وإذا كانت y_h وx_h مجاميع العيّنة في الطبقة h وX_h مجموع المقادير x_{hi} ضمن الطبقة h ، فإن التقدير $\hat{R}_{RS}$ (s ترمز لكون التقدير منفصلاً) هو

$$\hat{Y}_{Rs} = \sum_h \frac{y_h}{x_h} X_h = \sum_h \frac{\bar{y}_h}{\bar{x}_h} X_h \tag{6.44}$$

ولم نضع أية فروض حول بقاء النسب الحقيقية ثابتة من طبقة إلى أخرى. ويتطلب التقدير معرفة المجاميع المنفصلة X_h .

نظرية (٦-٤)

إذا سحبنا عيّنات عشوائية بسيطة مستقلة واحدة من كل طبقة وكانت حجوم هذه العيّنات n_h كبيرة في جميع الطبقات، فعندئذ

$$V(\hat{Y}_{Rs}) = \sum_h \frac{N_h^2(1-f_h)}{n_h}(S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h \rho_h S_{yh} S_{xh}) \tag{6.45}$$

حيث $R_h = Y_h/X_h$ هي النسبة الحقيقة في الطبقة h و ρ_h معرّفة كما سبق ضمن كل طبقة.

برهـان

بتطبيق العلاقة (6.2) ، فقرة ٦-٣، الخاصة بعيّنة عشوائية بسيطة نجد في الطبقة h ،

$$V(\hat{Y}_{Rh}) = \frac{N_h^2(1-f_h)}{n_h}(S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h \rho_h S_{yh} S_{xh}) \tag{6.46}$$

وبما أن $\hat{Y}_{Rs} = \sum_h \hat{Y}_{Rh}$ والمعاينة مستقلة في كل طبقة، فلدينا $V(\hat{Y}_{Rs}) = \sum V(\hat{Y}_{Rh})$ ومنه النتيجة المطلوبة في (6.45) .

ولا تكون هذه العلاقة مشروعة إلا إذا كانت العيّنة ضمن كل طبقة كبيرة إلى الحد الذي يسمح بتطبيق العلاقة التقريبية للتباين على كل طبقة. ويجب الانتباه إلى هذه الشروط في التطبيقات العملية.

وفضلاً عن ذلك، عندما تكون المقادير n_h صغيرة وعدد الطبقات L كبيراً، فقد لا يكون الانحياز في $\hat{Y}_{Rs}$ قابلاً للإهمال بالمقارنة مع خطئه المعياري، كما تقترح المحاكمة التقريبية التالية

رأينا في طبقة واحدة (فقرة ٦ ـ ٨) أن

$$\frac{|\text{الانحياز في } \hat{Y}_{Rh}|}{\sigma(\hat{Y}_{Rh})} \leq \text{cv of } \bar{x}_h \qquad \text{معامل اختلاف}$$

وإذا كان للانحياز الإشارة نفسها في جميع الطبقات، كما قد يحدث، فإن الانحياز في $\hat{Y}_{Rs}$ سيساوي تقريبًا L مرة الانحياز في $\hat{Y}_{Rs}$ ولكن الخطأ المعياري لِـ $\hat{Y}_{Rs}$ هو فقط من مرتبة $\sqrt{L}$ مضروبًا بالخطأ المعياري لِـ $\hat{Y}_{Rs}$ وهكذا تكون النسبة :

$$\frac{|\text{الانحياز في } \hat{Y}_{Rs}|}{\sigma(\hat{Y}_{Rs})}$$

من مرتبة

$$\sqrt{L}(\ \bar{x}_h \ \text{(معامل اختلاف})$$

وعلى سبيل المثال، مع وجود 50 من الطبقات ومع كون معامل اختلاف $\bar{x}_h$ حوالي 0.1 في كل طبقة، يمكن أن يكون الانحياز في $\hat{Y}_{Rs}$ مساويًا جداء خطئه المعياري بِـ 0.7 ومساهمة الانحياز في متوسط مربعات الخطأ لِـ $\hat{Y}_{Rs}$ ستبلغ حوالي الثلث.

ومع أن الانحياز يكون عمليًا أصغر في العادة بكثير من حدّه الأعلى، فإن خطورة الانحياز في التقدير النسبة المنفصل يجب أن تبقى ماثلة في الذهن إذا كان

(معامل اختلاف $\bar{x}_x$) $\sqrt{L}$ يتجاوز، مثلاً، 0.3 .

(٦-١١) التقدير النسبة المركب

يُستمد تقدير بديل من نسبة مركبة واحدة (Herwitz, Hansen و 1946, Gurney) فلنحسب من البيان الإحصائي للعيّنة :

$$\hat{Y}_{st} = \sum_h N_h \bar{y}_h, \qquad \hat{X}_{st} = \sum_h N_h \bar{x}_h \tag{6.47}$$

وهذان التقديران هما التقديران المعروفان لمجموعي المجتمعين Y و X على الترتيب، والناتجين عن عيّنة طبقية. والتقدير النسبة المركب $\hat{Y}_{Rc}$ (c لتدلّ على أن التقدير مركب) هو

$$\hat{Y}_{Rc} = \frac{\hat{Y}_{st}}{\hat{X}_{st}} X = \frac{\bar{y}_{st}}{\bar{x}_{st}} X \tag{6.48}$$

حيث $\bar{x}_{st} = \hat{X}_{st}/N$ ، $\bar{y}_{st} = \hat{Y}_{st}/N$ هما تقديرا متوسطي المجتمع المحسوبين من عيّنة طبقية.

ولا يتطلب التقدير $\hat{Y}_{Rc}$ ، معرفة بالمقادير X_h ولكن فقط بـ X .

والتقدير المركب أقل خضوعًا بكثير لمخاطرة الانحياز من التقدير المنفصل.

ومستخدمين طريقة Ross و Hartley في الفقرة (٦ ـ ٨)، نجد معتبرين $\hat{R}_c = \bar{y}_{st}/\bar{x}_{st}$،

$$\begin{aligned}\operatorname{cov}(\hat{R}_c, \bar{x}_{st}) &= E\left(\frac{\bar{y}_{st}}{\bar{x}_{st}} \cdot \bar{x}_{st}\right) - E(\hat{R}_c)E(\bar{x}_{st}) \\ &= \bar{Y} - \bar{X}E(\hat{R}_c)\end{aligned} \tag{6.49}$$

ومنه

$$E(\hat{R}_c) = R - \frac{1}{\bar{X}} \operatorname{cov}(\hat{R}_c, \bar{x}_{st})$$

و

$$\frac{|\text{الانحياز في } \hat{R}_c|}{\sigma_{\hat{R}_c}} = \frac{|\rho_{\hat{R}_c, \bar{x}_{st}} \cdot \sigma_{\bar{x}_{st}}|}{\bar{X}} \leq \text{معامل اختلاف } \bar{x}_{st} \tag{6.50}$$

وهكذا تكون الانحيازات في $\hat{R}_c$ ، $\hat{Y}_{Rc}$ قابلة للإهمال بالنسبة إلى أخطائها المعيارية، شريطة فقط أن يكون معامل اختلاف $\bar{x}_{st}$ أقل من 0.1 .

نظريـة (٦-٥)

إذا كان حجم العينة الكلي n كبيراً فإن

$$V(\hat{Y}_{Rc}) = \sum_h \frac{{N_h}^2(1-f_h)}{n_h}({S_{yh}}^2 + R^2{S_{xh}}^2 - 2R\rho_h S_{yh}S_{xh}) \qquad (6.51)$$

برهان

نتّبع هنا المناقشة نفسها التي اتبعناها في النظرية (٢-٥) . وفي حالتنا هنا نجد أن المعادلة الرئيسة هي

$$(\hat{Y}_{Rc} - Y) = \frac{N\bar{X}}{\bar{x}_{st}}(\bar{y}_{st} - R\bar{x}_{st}) \doteq N(\bar{y}_{st} - R\bar{x}_{st}) \qquad (6.52)$$

لنعتـبر الآن المتغـير $u_{hi} = y_{hi} - Rx_{hi}$ والـطرف الأيمن من المعـادلـة (6.52) هو $N\bar{u}_{st}$ حيث $\bar{u}_{st}$ هو المتوسط المرجّح للمتغير u_{hi} في عيّنة طبقية . بالإضافة إلى أن متوسط المجتمع $\bar{U}$ للمقادير u_{hi} هو الصفر، باعتبار أن $R = \bar{Y}/\bar{X}$. وهكذا يمكن تطبيق النظرية (٥-٣) المتعلقة بتباين تقدير متوسط من عيّنة عشوائية طبقية ، على $\bar{u}_{st}$ وهذا يعطي

$$V(\hat{Y}_{Rc}) = N^2 V(\bar{u}_{st}) = \sum_h \frac{N_h(N_h - n_h)}{n_h}{S_{uh}}^2 \qquad (6.53)$$

حيث

$$\begin{aligned}{S_{uh}}^2 &= \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(u_{hi} - \bar{U}_h)^2 \\ &= \frac{1}{N_h - 1}\sum_{i=1}^{N_h}[(y_{hi} - \bar{Y}_h) - R(x_{hi} - \bar{X}_h)]^2\end{aligned}$$

ونحصل على النتيجة (6.51) عند فكّ المربع .

ومن المفيد أن نلاحظ من المعـادلتـين (6.45) و (6.51) أن التبـاينين التقريبيين لِـ $\hat{Y}_{Rs}$ و $\hat{Y}_{Rc}$ يفترضان الشكل العام نفسه، والفرق الوحيد هو أن نسب المجتمع R_h في كل طبقة بمفردها والموجودة في (6.45) يحل محلها R في (6.51)

(٦-١٢) مقارنة التقديرين المركب والمنفصل

يمكن كتابة

$$V(\hat{Y}_{Rc})-V(\hat{Y}_{Rs})$$
$$=\sum_h \frac{N_h^2(1-f_h)}{n_h}[(R^2-R_h^2)S_{xh}^2-2(R-R_h)\rho_h S_{yh}S_{xh}]$$
$$=\sum \frac{N_h^2(1-f_h)}{n_h}[(R-R_h)^2S_{xh}^2+2(R_h-R)(\rho_h S_{yh}S_{xh}-R_hS_{xh}^2)]$$

وعادة يكون الحد الأخير من الطرف الأيمن صغيراً في الحالات التي يكون فيها التقـدير النسـبة ملائمًا، (وينعدم إذا كانت العلاقة بين x_{hi} وy_{hi} ضمن كل طبقة هي علاقة خط مستقيم يمر من المبدأ). وهكذا فإنه من المحتمل أن يكون استخدام تقدير منفصل ضمن كل طبقة أكثر دقة، ما لم يبق R_h ثابتًا من طبقة إلى أخرى، هذا إذا كانت العيّنـة ضمن كل طبقـة كبـيرة بكفاية بحيث تصبح العلاقة التقريبية لِـ $V(\hat{Y}_{Rs})$ مشروعـه، وكان الانحياز المتجمع الذي يمكن أن يؤثر في $\hat{Y}_{Rs}$ (فقرة ٦-١٠) قابلاً للإهمـال. وفي حالـة عيّنـة صغيرة فقط ضمن كل طبقة، يُوصى باستخدام التقدير المركب ما لم توجد دلالة تجريبية قوية تفيد العكس.

وللحصول على تقديرات العيّنة لهذه التباينات نبدل تقديرَي R_h وR المحسوبَين من العيّنة في الأمكنة المناسبة كما نعوّض متوسطات مربعات العيّنة s_{yh}^2 و s_{xh}^2 من أجل التباينات الموافقة، وتغاير العيّنة من أجل الحد $\rho_h S_{yh}S_{xh}$. ويجب حساب كل من متوسط مربعات العيّنة وتغاير العيّنة بصورة منفصلة في كل طبقة.

مثـــال

البيان الإحصائي ناشىء عن تعداد إحصائي لكل المزارع في منطقة جيفرسون، ولاية أيوا. وتمثل y_{hi} في هذا المثال عدد فدادين الذرة، وx_{hi} عدد الفدادين في المزرعة. وقـد قُسم المجتمـع إلى طبقتين، تحوي الطبقة الأولى مزارع يصل حجمها إلى 160 فدانًا. ونأخذ عيّنة حجمها 100 مزرعة. وسنفترض عند استخدام المعاينة الطبقية أننا

أخـذنـا 70 مزرعة من الطبقة الأولى و30 مزرعة من الطبقة الثانية، وهي، على وجه التقريب، المحاصّة المثلى. والبيان الإحصائي معطى في الجدول (٦-٣). والكميات الثـلاث الأخـيرة Q_h, V_h'، V_h'' هي كميات مسـاعـدة سنستخـدمها في الحسابات، والكميتان الأخيرتان معرفتان فيما بعد.

ونعتـبر خمس طرق لتقـدير متوسط المجتمع لعدد فدادين المزرعة الواحدة من الذرة. وسنتجاهل عامل الـ ت م م.

١ - عيّنة عشوائية بسيطة: تقدير المتوسط للمزرعة الواحدة.

$$V_1 = \frac{S_y^2}{n} = \frac{620}{100} = 6.20$$

جدول (٦-٣) البيان الإحصائي من منطقة جيفرسون، أيوا

الطبقات	حجم المزرعة بالفدادين	N_h	S_{yh}^2	S_{yxh}	S_{xh}^2	R_h
1	0–160	1580	312	494	2055	0.2350
2	More than 160	430	922	858	7357	0.2109
للمجتمع بكامله		2010	620	1453	7619	0.2242

الطبقات	$\bar{Y}_h$	$\bar{X}_h$	n_h	$Q_h = W_h^2/n_h$	V_h'	V_h''
1	19.40	82.56	70	0.008828	193	194
2	51.63	244.85	30	0.001525	887	907
للمجتمع بكامله	26.30	117.28	100			

٢ - عيّنة عشوائية بسيطة: التقدير النسبة

$$V_2 - \frac{1}{n}(S_y^2 + R^2 S_x^2 - 2RS_{yx})$$

$$= \frac{1}{100}[620 + (0.2242)^2(7619) - 2(0.2242)(1453)]$$

$$= 3.51$$

٣ - عيّنة عشوائية طبقية: تقدير متوسط المزرعة الواحدة

$$V_3 = \sum \frac{W_h^2}{n_h} S_{yh}^2 = \sum Q_h S_{yh}^2 = 4.16$$

٤ ـ عيّنة عشوائية طبقية: التقدير النسبة مستخدمين نسبة منفصلة في كل طبقة.

$$V_4 = \sum Q_h (S_{yh}{}^2 + R_h{}^2 S_{xh}{}^2 - 2R_h S_{yxh}) = \sum Q_h V_h' = 3.06$$

٥ ـ معاينة عشوائية طبقية: التقدير النسبة مستخدمين نسبة مركبة.

$$V_5 = \sum Q_h (S_{yh}{}^2 + R^2 S_{xh}{}^2 - 2RS_{yxh}) = \sum Q_h V_h'' = 3.10$$

ويمكن تلخيص الدقة النسبية للطرق المختلفة كما يلي:

طريقة المعاينة	طريقة التقدير	الدقة النسبية
١ ـ عشوائية بسيطة	المتوسط للمزرعة الواحدة	100
٢ ـ عشوائية بسيطة	النسبة	177
٣ ـ عشوائية طبقية	المتوسط للمزرعة الواحدة	149
٤ ـ عشوائية طبقية	نسبة منفصلة	203
٥ ـ عشوائية طبقية	نسبة مركبة	200

وتُبرز النتائج نقطة مهمة لها تطبيقات واسعة. فالتقسيم إلى طبقات وفقًا لحجم المزرعة يحقق الغرض العام نفسه الذي يحققه تقدير نسبة مقامه هو حجم االمزرعة. والـطريقتـان تخفضان تأثير تغيرات حجم المزرعة في خطأ المعاينة عند تقدير متوسط فدادين الذرة للمزرعة الواحدة. وعلى سبيل المثال، فإن الكسب في الدقة من التقدير النسبـة هو 77 بالمـائـة عنـد استخـدام المعـاينة العشوائية البسيطة، ولكنه 36 بالمائة فقط (203 مقابل 149) عند استخدام معاينة عشوائية طبقية.

وعند تصميم مسح عيّنة قد يكون لنا الخيار بين أن نعتمد عاملاً ما في عملية تحديد الطبقات أو نستخدمه في طريقة التقدير. ويعتمد أفضل قرار على الظروف. ومن المفيد هنا ذكر النقاط التالية: (ا) إن اعتماد بعض العوامل، كالموقع الجغرافي مثلاً، في تحديد الطبقات أيسر من استخدامه في طريقة التقدير. (ب) وتعتمد المسألة على طبيعة العلاقة بين x_i وy_i وتعمل جميع الطرق البسيطة في التقدير بأفضل فعالية عند وجود علاقة خطيّة. وفي حال وجود علاقة معقدة أو غير مستمرة فقد يكون التقسيم إلى طبقات أكثر فعالية، ذلك لأن التقسيم إلى طبقات، في حال وجود عدد كاف منها، سيلغي تأثيرات

العلاقة بين x_i وy_i أيًا كان نوعها تقريبًا. (جـ) وإذا كانت بعض المتغيرات المهمة متناسبة تقريبًا مع x_i ، إلا أن متغيرات أخرى تتناسب تقريبًا مع متغير آخر z_i فمن الأفضل استخدام x_i وz_i كمقامات في التقديرات النسبة بدلاً من استخدام أحدها في التقسيم إلى طبقات.

(٦-١٣) طريقة مختزلة لحساب تقدير تباين

إذا كان $n_h=2$ في جميع الطبقات، فقد أعطى Keyfitz (1957) طرقًا بسيطة لحساب تقريبات للتباينات المقدّرة لـ $\hat{Y}_{Rc}$ أو $\hat{R}_c$ أو بصورة أعم لدوالّ في متغير أو أكثر من الشكل $\hat{Y}_{st}$ ومن أجل $\hat{R}_c$ لدينا:

$$\hat{R}_c=\frac{\hat{Y}_{st}}{\hat{X}_{st}}=\frac{\sum_h \hat{Y}_h}{\sum_h \hat{X}_h}=\frac{\sum_h \frac{N_h}{2}(y_{h1}+y_{h2})}{\sum_h \frac{N_h}{2}(x_{h1}+x_{h2})} \qquad (6.54)$$

وتستخدم طريقة Keyfitz المطابقة التالية من أجل $n_h=2$ ،

$$2s_{yh}{}^2=2\sum_{i=1}^{2}(y_{hi}-\bar{y}_h)^2\equiv(y_{h1}-y_{h2})^2=(dy_h)^2 \qquad (6.55)$$

حيث $dy_h=(y_{hi}-y_{h2})$ ، وبالتالي

$$v(\hat{Y}_h)=\left(\frac{N_h}{2}\right)^2 2(1-f_h)s_{yh}{}^2=(1-f_h)(y_{h1}'-y_{h2}')^2=(1-f_h)(dy_h')^2 \qquad (6.56)$$

حيث $y_{hi}'=N_h y_{hi}/2$ وبصورة مماثلة، من أجل تقدير العيّنة للتغاير.

$$\text{cov}(\hat{Y}_h\hat{X}_h)=(1-f_h)(dy_h')(dx_h') \qquad (6.57)$$

والآن

$$\hat{R}_c-R=\frac{\hat{Y}_{st}}{\hat{X}_{st}}-R\doteq\frac{\hat{Y}_{st}-R\hat{X}_{st}}{X}=\frac{Y}{X}\left(\frac{\hat{Y}_{st}}{Y}-\frac{\hat{X}_{st}}{X}\right) \qquad (6.58)$$

وبما أن المعاينة مستقلة في الطبقات المختلفة فنجد،

$$v(\hat{R}_c)=\left(\frac{Y}{X}\right)^2\sum_h(1-f_h)\left[\left(\frac{dy_h'}{Y}\right)^2+\left(\frac{dx_h'}{X}\right)^2-2\left(\frac{dy_h'}{Y}\right)\left(\frac{dx_h'}{X}\right)\right] \qquad (6.59)$$

$$= \left(\frac{Y}{X}\right)^2 \sum_h (1-f_h) \left(\frac{dy_h'}{Y} - \frac{dx_h'}{X}\right)^2 \tag{6.60}$$

وقد مدّ Keyfitz هذه الطريقة لتغطي مقدّرات ما بعد التقسيم إلى طبقات والمعاينة متعددة المراحل، ولتعطي تباينات الفروق بين التقديرات من مسوح متتالية في نظام عينات دورية. ويعطي Woodruff (1971) معالجة عامة تتناول التقديرات غير الخطية، واحتمالات اختيار غير متساوية، وعيّنات حجمها n_h في كل طبقة. وتوضيحًا لطريقة Woodruff لنعتبر دالة $f(\hat{\mathbf{Y}})$ حيث يمثل $\hat{Y}$ المتجه أو المجموعة من m من المتغيرات $\hat{Y}_j = \sum_h \hat{Y}_{jh}$. ومع معاينة عشوائية بسيطة في الطبقة h نجد أن الـ $\hat{Y}_{jh}$ هي من الشكل $(N_h/n_h)\sum y_{jhi}$، ويمتدّ المجموع فوق وحدات المعاينة في الطبقة h وعددها n_h. ووفقًا لتقريب تايلور نجد،

$$f(\hat{\mathbf{Y}}) - f(\mathbf{Y}) \doteq \sum_j^m \frac{\partial f}{\partial Y_j}(\hat{Y}_j - Y_j) = \sum_j \sum_h \frac{\partial f}{\partial Y_j}(\hat{Y}_{jh} - Y_{jh}) \tag{6.61}$$

والحيلة هنا هي عكس ترتيب إشارتي المجموع بحيث نكتب،

$$f(\hat{\mathbf{Y}}) - f(\mathbf{Y}) \doteq \sum_h \sum_j \frac{\partial f}{\partial Y_j}(\hat{Y}_{jh} - Y_{jh}) = \sum_h (\hat{U}_h - U_h) \tag{6.62}$$

حيث

$$\hat{U}_h = \sum_j \frac{\partial f}{\partial Y_j}\hat{Y}_{jh} = \frac{N_h}{n_h}\sum_i^{n_h}\left(\sum_j \frac{\partial f}{\partial Y_j} y_{jhi}\right) = \sum_i^{n_h} u_{hi} \tag{6.63}$$

وبحساب المشتقات $\frac{\partial f}{\partial Y_j}$ عند التقديرات $\hat{Y}_j$، يمكن حساب الـ u_{hi} من (6.63) لكل وحدة معاينة في الطبقة h. ومع معاينة عشوائية بسيطة في الطبقات، تنطبق العلاقة المعتادة الخاصة بتقدير تباين المجموع $\hat{Y}_h$ وبالتالي نجد من (6.62) تقدير عيّنة تقريبيًا لتباين $f(\hat{Y})$ هو

$$v[f(\hat{\mathbf{Y}})] \doteq \sum_h \frac{(N_h - n_h)n_h}{N_h} \frac{\sum_i^{n_h}(u_{hi} - \bar{u}_h)^2}{(n_h - 1)} \tag{6.64}$$

وميزة هذه الطريقة هي أننا لا نحتاج إلى حساب تغايرات المقادير $\hat{Y}_{jh}$.

ولإيضاح الـ u_{hi} في حالة المقدّر $\hat{R}_c = \hat{Y}/\hat{X}$ ستكون كتابة $\hat{R}_c = \hat{Y}_1/\hat{Y}_2$ مفيدة، بحيث نجد،

$$f(\mathbf{Y}) = \frac{Y_1}{Y_2}: \quad \frac{\partial f}{\partial Y_2} = \frac{1}{Y_2}: \quad \frac{\partial f}{\partial Y_2} = -\frac{Y_1}{Y_2^2} \tag{6.65}$$

وبالتالي نجد من (6.63) في حالة $n_h=2$

$$u_{hi} = \frac{N_h}{n_h}\left(\frac{y_{1hi}}{\hat{Y}_2} - \frac{\hat{Y}_1}{\hat{Y}_2^2} y_{2hi}\right) = \frac{N_h}{n_h} \frac{(y_{1hi} - \hat{R}_c y_{2hi})}{\hat{Y}_2} \tag{6.66}$$

وبدلالة الـ y_{hi} ، x_{hi} الأصلية نجد،

$$u_{hi} = \frac{N_h}{n_h} \frac{(y_{hi} - \hat{R}_c x_{hi})}{\hat{X}} \tag{6.67}$$

وتقدير تباينها معطى على وجه التقريب في (6.64) .

وقد استُخدمت طريقة Keyfitz ، على سبيل المثال، في مسح مقابلة صحي قام به المركز القومي للإحصاءات الصحية. ووحدة المعاينة هي وحدة عنقودية ـ منطقة أو مجموعة مناطق متجاورة. وقُسمت كل وحدة إلى قطاعات يتضمن الواحد مِنها حوالي 9 أُسر، وقد اختير في المتوسط 13 قطاعًا من كل وحدة خضعت للمعاينة. ويمكن أن يكون المتغيران y_{h1} ، y_{h2} أعداد الأشخاص المصابين بمرض معيّن و x_{h1} ، x_{h2} الأعداد الكلية للأشخاص في العيّنات المأخوذة من وحدتين من الطبقة h .

وبالإضافة إلى التقسيم الجغرافي الابتدائي إلى طبقات من المناطق هناك تقسيم لاحق لأشخاص العيّنة إلى طبقات وفقًا للعمر، الجنس، واللون. وهكذا، وبدلاً من المقدّر $\hat{Y}_{Rc} = X\hat{R}_c$ لعدد المرضى الكلي يكون المقدّر

$$\hat{Y}_{PS} = \sum_a X_a \hat{R}_a = \sum_a X_a \frac{\hat{Y}_a}{\hat{X}_a} \tag{6.68}$$

حيث يمثل a فئة عمر ـ جنس ـ لون و X_a المجموع المعروف للمجتمع ضمن هذه الفئة. ولدينا هنا دالّة بمجموعتين من المتغيرات العشوائية $\hat{Y}_a$ و $\hat{X}_a$. وفضلاً عن ذلك، ولعدة أسباب، قد يوجد ارتباط بين y_{ahi} و $y_{a'hi}$ من أجل فئتين مختلفين a', a ضمن وحدة

عنقودية ؛ وعلى سبيل المثال، في حالة مرض معدٍ قد يكون عدد الحالات مرتفعًا في جميع فئات الوحدة. وبتطبيق (6.63) متجاهلين الـ ت م م نجد في حالة $n_h=2$ ،

$$v(\hat{Y}_{PS})=\sum_h\left[\sum_a X_a\left(\frac{dy_{ah}'}{\hat{X}_a}-\frac{\hat{Y}_a dx_{ah}'}{\hat{X}_a^2}\right)\right]^2 \tag{6.69}$$

$$\doteq\sum_h\left[\sum_a X_a\left(\frac{\hat{Y}_a}{\hat{X}_a}\right)\left(\frac{dy_{ah}'}{\hat{Y}_a}-\frac{dx_{ah}'}{\hat{X}_a}\right)\right]^2 \tag{6.70}$$

حيث $dy_{ah}'=N_h(y_{ah1}-y_{ah2})/2$.

وفي هذا التطبيق تتناول طريقة Keyfitz أيضًا تعقيدات إضافية في المسح الإحصائي - اختيار الوحدات الابتدائية باحتمالات غير متساوية، والتعديلات في حالات عدم الاستجابة، واستخدام طريقة الطبقة المنهارة. وفضلاً عن ذلك، وباعتبار أن وقت الحاسب حتى مع هذه الطريقة البسيطة لا يسمح بحسابات التباين إلا لعدد محدود من المفردات فقط، فقد أُعطيت جداول للعلاقة بين $\frac{\text{الخطأ المعياري } (\hat{Y})}{\hat{Y}}$ و $\hat{Y}$

وذلك في أنواع مختلفة من المفردات كي تساعد في التنبؤ بالخطأ المعياري لِـ $(\hat{Y})$ في مفردات لم نحسب خطأها المعياري. ويعطي Bean (1970) عرضًا واضحًا لهذه الطرق والنتائج.

(٦-١٤) المحاصّة المثلى في حالة التقدير النسبة

يمكن أن تختلف المحاصّة المثلى لِـ n_h عند استخدام التقدير النسبة عنها عند استخدام المتوسط لكل وحدة. لنعتبر أولاً المتغير $\hat{Y}_{Rs}$ فمن النظرية (٦-٤) نجد أن تباينه :

$$V(\hat{Y}_{Rs})=\sum_h\frac{N_h(N_h-n_h)}{n_h}(S_{yh}^2+R_h^2S_{xh}^2-2R_h\rho_hS_{yh}S_{xh})$$

$$=\sum_h\frac{N_h(N_h-n_h)}{n_h}S_{dh}^2,\quad \text{with } S_{dh}^2=\frac{1}{N_h-1}\sum_{i=1}^{N_h}d_{hi}^2 \tag{6.71}$$

حيث $d_{hi}=y_{hi}-R_hx_{hi}$ هو انحراف y_{hi} عن R_hx_{hi} . ومن الطرق المعطاة في الفصل الخامس لإيجاد محاصّة مثلى نجد أن (6.71) ستكون، تحت شرط تكلفة إجمالية من

الشكل $\sum c_h n_h$ أصغر ما يمكن، عندما يكون

$$n_h \propto \frac{N_h S_{dh}}{\sqrt{c_h}}$$

وكما نذكر فإننا في حالة المتوسط لكل وحدة، ومن أجل تباين أصغري، نختار n_h بحيث يكون متناسبًا مع $N_h S_{yh}/\sqrt{c_h}$.

وقد تبدو المحاصّة في حالة التقدير النسبة محيّرة قليلاً عند تخطيط العيّنة. إذ يبدو من الصعب تخمين القيم الممكنة لـ S_{dh} . وهناك قاعدتان مفيدتان. ففي مجتمع يكون فيه التقـدير النسبـة أفضل تقدير خطي غير منحاز، سيكون S_{dh} متناسبًا، على وجه التقـريب، مع $\sqrt{\bar{X}_h}$ (بـالاستناد إلى النظرية ٦-٣) وفي هذه الحالة ينبغي أن يكون n_h متنـاسبًـا مع $N_h\sqrt{\bar{X}_h}/\sqrt{c_h}$ وأحيانًا يمكن أن يكون تباين d_{hi} أقرب إلى وضع التنـاسب مع $\bar{X}_h^2$ وهـذا يقـود إلى محاصّـة لـ n_h متناسبة مع $N_h\bar{X}_h/\sqrt{c_h}$, أي مجموع الـطبقـة X_{hi} مقسـومًـا على الجـذر الـتربيعي لتكلفـة الـوحـدة. وقـد ناقش Herwitz, Hansen و Gurney (1946) مثـالاً من هذا النـوع، وذلك في عيّنة مصمّمة لتقدير مبيعات مخازن المفرّق.

وتنطبق المناقشة العامة نفسها إذا كنا سنستخدم التقدير $\hat{Y}_{RC}$.

مثـال

يمكن مقـارنة الطرق المختلفة للتحاصّ من بيان إحصائي جُمع نتيجة لتعداد تام لـ 257 من بسـاتـين الـدرّاق في ولاية نورث كارولاينـا في حزيران 1946 (Finkner, 1950) . والغاية من هذا المسح الإحصائي هو تحديد طريقة المعاينة الأكثر كفاءة لتقدير الإنتاج التجاري للدرّاق في هذه المنطقة. وقد جُمعت المعلومات حول عدد أشجار الدرّاق والإنتاج التقديري الكلي من الدرّاق في كل بستان. والارتباط العالي بين هذين المتغيرين اقترح استخدام التقدير النسبة. وقد أُلغي بستان كبير جدًا.

وفي هذا التوضيح، قُسمت المساحة جغرافيًا إلى ثلاث طبقات، وقد رمزنا لعدد أشجـار الـدرّاق في بستـان بـ x_{hi} ولـلإنتـاج التقـديري مقـاسًـا بالبوشل من الدرّاق

بـ y_{hi} . وسنعتبر فقط التقدير النسبة الأول $\hat{Y}_{Rs}$ (مبنيًّا على نسبة منفصلة في كل طبقة)، باعتبار أن المبدأ يبقى نفسه من أجل النوعين كليهما، التقدير الطبقي والتقدير النسبة .

وسنقارن أربع طرق مختلفة للتحاصّ: (ا) متناسب مع N_h ، (ب) n_h متناسب مع N_hS_{yh} ، (جـ) n_h متنـــاسب مع $N_h\sqrt{\bar{X}_h}$، و(د) n_h متنـــاسب مع $N_h\bar{X}_h = X_h$. وسنفترض حجمًا مساويًا لِـ 100 ويلخص الجدول (٦-٤) البيان الإحصائي لهذه المقارنات .

جدول (٦-٤) بيان عن المسح الإحصائي للدراق في نورث كارولاينا

الطبقات	S_{xh}^2	S_{yxh}	S_{yh}^2	S_{xh}	S_{yh}	$\bar{X}_h$	$\bar{Y}_h$	R_h	S_{dh}^2
1	5186	6462	8699	72.01	93.27	53.80	69.48	1.29133	658
2	2367	3100	4614	48.65	67.93	31.07	43.64	1.40475	573
3	4877	4817	7311	69.83	85.51	56.97	66.39	1.16547	2706
المجتمع	3898	4434	6409	62.43	80.06	44.45	56.47	1.27053	1433

الطبقات	N_h	(a)	N_hS_{yh}	(b)	$\sqrt{\bar{X}_h}$	$N_h\sqrt{\bar{X}_h}$	(c)	$N_h\bar{X}_h$	(d)
1	47	18	4384	22	7.33	344.5	20	2529	22
2	118	46	8016	40	5.57	657.3	39	3666	32
3	91	36	7781	38	7.55	687.1	41	5184	46
المجتمع	256	100	20181	100	20.45	1688.9	100	11379	100

ويبـيّن الجـزء العلوي من الجـدول المعلومات الأساسية . والطريقة التي استُخدمت لحساب التباينات الأربعة هي أن نجد أولاً المقادير n_h لكل نوع من المحاصّة . وهذه القيم مبينة في الأعمدة (a) إلى (d) في النصف الأسفل من الجدول . وهكذا فإنه مع المحاصّة (ا) نجد $n_h = nN_h/N,$ أي أنه في المحاصّة الأولى يكون

$$n_1 = \frac{(100)(47)}{256} = 18$$

وبعد الحصول على قيم n_h ، نحسب $V(\hat{Y}_{Rs})$ الموافق بالتعويض في العلاقة :

$$V(\hat{Y}_{Rs}) = \sum_h \frac{N_h(N_h - n_h)}{n_h} S_{dh}^2$$

حيث

$$S_{dh}^2 = S_{yh}^2 + R_h^2 S_{xh}^2 - 2R_h S_{yxh}$$

والكميات s_{xh}^2 معطاة في أقصى اليمين من النصف الأعلى للجدول (٦-٤).

جدول (٦-٥) مقارنة أربع طرق في التحاصّ

طريقة التحاصّ: متناسبة مع n_h	التباين: الطبقات 1	2	3	Total	Relative Precision
1. N_h	49,824	105,833	376,215	531,872	100
2. $N_h S_{yh}$	35,144	131,847	343,446	510,437	104
3. $N_h \sqrt{\bar{X}_h}$	41,750	136,964	300,312	479,026	111
4. $N_h \bar{X}_h$	35,144	181,710	240,888	457,742	116

ويبينّ الجدول (٦-٥) التباينات والدقة النسبية.

ولا يوجد الكثير من الاختيار بين المحاصّات المختلفة، كما قد نتوقع، باعتبار أن n_h لا يختلف كثيراً في الطرق الأربع. وتُظهر الطريقة 4 حيث المحاصّة متناسبة مع العدد الكلي لأشجار الدراق في الطبقة، تفوقًا ضئيلاً على الطرق الأخرى.

(٦-١٥) التقديرات النسبة غير المنحازة

وإذا بدا لنا أن التقدير النسبة المنفصل هو التقدير المناسب، فقد تكون التقديرات النسبة غير المنحازة، أو التي يقل انحيازها عن انحياز $\hat{R}$ و $\hat{Y}_R$ تقديرات مفيدة، كما لاحظنا، في مسوح تتضمن العديد من الطبقات مع عينة صغيرة من كل طبقة. وتوجد ثلاث طرق تعطي تقديرات غير منحازة وثلاث طرق لإزالة الحدّ من مرتبة $\frac{1}{n}$ في الانحياز [انظر (6.33)]. وسنناقشها هنا باختصار.

عند مقارنة هذه الطرق تطرح الأسئلة التالية نفسها: (ا) هل تنافس قيمة متوسط مربعات الخطأ في هذه الطريقة تلك المتعلقة بالتقدير النسبة العادي؟ (ب) هل تقدم الطريقة تقدير عيّنة مرضيا للتباين؟ وهذه تمثل صعوبة بالنسبة لـ $\hat{R}$ كما رأينا. وسنصف في البداية هذه الطرق. وتتطلب الطرق غير المنحازة معرفة $\bar{X}$.

طرق غير منحازة

يمكن استنتاج أحد التقديرات، ويعود إلى Hartley و Ross (1954) بالبدء

بمتوسط النسب y_i/x_i وليكن $\bar{r}$ ثم تصحيحه من أجل الانحياز.

$$\bar{r} = \frac{1}{n}\sum^{n} r_i = \frac{1}{n}\sum^{n} \frac{y_i}{x_i}$$

والآن

$$\frac{1}{N}\sum_{i=1}^{N} r_i(x_i - \bar{X}) = \frac{1}{N}\sum_{i=1}^{N} \frac{y_i}{x_i} \cdot x_i - \left(\frac{1}{N}\sum_{i=1}^{N} r_i\right)\bar{X}$$

$$= \bar{Y} - \bar{X}E(r_i) = \bar{X}[R - E(r_i)] \tag{6.72}$$

ولكن $E(\bar{r}) = E(r_i)$ في معاينة عشوائية بسيطة. وبالتالي،

$$\text{الانحياز في } \bar{r} = E(\bar{r}) - R = -\frac{1}{\bar{X}N}\sum_{i=1}^{N} r_i(x_i - \bar{X}) \tag{6.73}$$

وبالاستناد إلى النظرية (٢-٣)، يكون

$$\frac{1}{n-1}\sum_{i=1}^{n} r_i(x_i - \bar{x}) = \frac{n}{n-1}(\bar{y} - \bar{r}\bar{x})$$

تقدير عيّنة غير منحاز لِـ

$$\frac{1}{N-1}\sum_{i=1}^{N} r_i(x_i - \bar{X})$$

وبالتعويض في (6.73) يصبح التقدير $\bar{r}$ بعد تصحيحه من أجل الانحياز:

$$\hat{R}_{HR} = \bar{r} + \frac{n(N-1)}{(n-1)N\bar{X}}(\bar{y} - \bar{r}\bar{x}) \tag{6.74}$$

والتقدير غير المنحاز الموافق لمجموع المجتمع $\hat{Y}$ هو

$$\hat{R}_{HR}X = \bar{r}X + \frac{n(N-1)}{n-1}(\bar{y} - \bar{r}\bar{x}) \tag{6.75}$$

وبمناقشات مماثلة، نستنتج تقديرًا غير منحاز آخر (Mickey, 1959) من النسب R_i والتي نحصل عليها بأن نحذف من العيّنة كل وحدة من وحداتها الـ n على التوالي، وفي كل مرة نحسب $\hat{R}_i = \sum y / \sum x$ من العناصر الـ $n-1$ الباقية. وإذا رمزنا بِـ $\hat{\bar{R}}$ لمتوسط الـ $\hat{R}_i$ هذه، فتقدير Mickey هو

$$\hat{R}_M = \hat{\bar{R}}_- + \frac{n(N-n+1)}{N\bar{X}}(\bar{y} - \hat{\bar{R}}_- \bar{x}). \qquad (6.76)$$

وكطريقة ثالثة، بينّ Lahiri (1951) أن التقدير النسبة العادي $\hat{R}$ غير منحاز، إذا كانت العيّنـة مسحـوبـة باحتمال متناسب مع $\sum x_i$. وربما كانت أبسط طريقة للقيام بهذا (Midzuna, (1951) هي سحب العنصر الأول من العيّنـة باحتـمال متناسب مع x_i. ثم سحب العناصر الـ $(n\text{-}1)$ الباقية من العيّنة باحتمالات متساوية. ومن السهل البرهان (تمرين ٦ـ١٠) على أن احتمال سحب عيّنة محددة بهذه الطريقة متناسب مع $\sum x_i$، وأن $\hat{R} = \sum^n y_i / \sum^n x_i$ غير منحاز في طريقة اختيار العيّنة هذه.

طرق بانحياز من مرتبة $1/n^2$

وتتألف هذه الطرق من تعديل لـ $\hat{R}$. وتعود أولاها إلى Quenouille (1956) وهي قابلة للتطبيق على صف واسع من المسائل الإحصائية يكون للتقدير المقترح فيها انحياز من مرتبة $\frac{1}{n}$. وقد أعطيت اسم «طريقة مدية الجيب» للدلالة على أداة لها العديد من الاستخدامات. وقد أشار Durbin (1959) إلى فائدة هذه الطريقة في التقديرات النسبة.

ومتجاهلين مؤقتًا الـ ت م م، يمكن نشر تقديرات منحازة مثل $\hat{R}$ في سلسلة من الشكل،

$$E(\hat{R}) = R + \frac{b_1}{n} + \frac{b_2}{n^2} + \cdots \qquad (6.77)$$

وإذا كان $n=mg$، لنقسم العيّنـة عشـوائيًا إلى g من الفئـات حجم كل منهـا m. ومن (6.77) نجد أن،

$$E(g\hat{R}) = gR + \frac{b_1}{m} + \frac{b_2}{gm^2} + \cdots \qquad (6.78)$$

والآن لتكن $\hat{R}_j$ النسبة المعتادة $\sum y / \sum x$ محسوبة من العيّنة بعد حذف الفئة j. وبما أننا نحصل على $\hat{R}_j$ من عيّنة عشوائية بسيطة حجمها $m(g\text{-}1)$، فلدينا

$$E(\hat{R}_j) = R + \frac{b_1}{(g-1)m} + \frac{b_2}{(g-1)^2 m^2} + \cdots \qquad (6.79)$$

وبالتالي

$$E[(g-1)\hat{R}_j] = (g-1)R + \frac{b_1}{m} + \frac{b_2}{(g-1)m^2} + \cdots \quad (6.80)$$

والطرح من (6.78) يعطي، إلى مرتبة n^{-2}،

$$E[g\hat{R} - (g-1)\hat{R}_j] = R - \frac{b_2}{g(g-1)m^2} = R - \frac{b_2}{n^2}\frac{g}{(g-1)}$$

والانحياز هو الآن من مرتبة $1/n^2$ ويمكن القيام بـ g من التقديرات من هذا النوع، واحد من أجل كل فئة. وتقدير Quenouille (مدية الجيب) هو متوسط هذه التقديرات g أي أن

$$\hat{R}_Q = g\hat{R} - (g-1)\hat{R}_- \quad (6.81)$$

حيث $\hat{R}_-$ هو متوسط المقادير $\hat{R}_j$ وعددها g . وكما بيّن Quenouille يختلف تباين $\hat{R}_Q$ عن تباين $\hat{R}$ بحدود من مرتبة $1/n^2$. وأية زيادة في التباين ناتجة عن مثل هذا التعديل من أجل الانحياز ينبغي أن تكون إذًا قابلة للإهمال في عيّنات معتدلة الحجم. ويبدو اختيار $g=n,\ m=1$ أفضل اختيار من أجل (مدية الجيب) في العيّنات الصغيرة.

وإذا لم يكن تجاهل الـ (ت م م) ممكنًا فإن الحد الرئيس في انحياز $\hat{R}$ هو، كما في (6.33)، من الشكل $b_1(1-f)/n$ ويمكن تبيان (تمرين ٦-١٠) أنه كي نزيل الحد من مرتبة $\frac{1}{n}$ والحد من مرتبة $\frac{1}{N}$ نحتاج إلى

$$\hat{R}_Q = w\hat{R} - (w-1)\hat{R}_- \quad (6.82)$$

حيث $w=g[1-(n-m)/N]$، أو في حالة $m=1$، $g=n$، $w=n[1-(n-1)/N]$.

ومقدّر Beal (1962) هو،

$$\hat{R}_B = \frac{\bar{y} + [(1-f)/n](s_{yx}/\bar{x})}{\bar{x} + [(1-f)/n](s_x{}^2/\bar{x})} = \hat{R}\cdot\frac{1+[(1-f)/n]c_{yx}}{1+[(1-f)/n]c_{xx}} \quad (6.83)$$

بينما نجد مقدّر Tin (1965) على الشكل،

$$\hat{R}_T = \hat{R}\left[1 - \frac{(1-f)}{n}\left(\frac{s_x{}^2}{\bar{x}^2} - \frac{s_{yx}}{\bar{y}\bar{x}}\right)\right] = \hat{R}\left[1 - \frac{(1-f)}{n}(c_{xx} - c_{yx})\right] \quad (6.84)$$

حيث $s_{yx}=\sum^n(y_i-\bar{y})(x_i-\bar{x})/(n-1)$ و $s_x{}^2=\sum^n(x_i-\bar{x})^2/(n-1)$ ، وهو على صلة وثيقـة بالمقدّر السابق. وهكذا يكون C_{yx} و C_{xx} التغـاير النسبي لِـ x و y والتباين النسبي لِـ x على الترتيب.

ويمكن إدراك بُنية $\hat{R}_T$ بأن نلاحظ من (6.34) إمكانية كتابة الحد الرئيس للقيمة المتوقعة لِـ $\hat{R}$ على الشكل،

$$R\left[1+\frac{(1-f)}{n}(C_{xx}-C_{yx})\right]$$

ومن الواضح أن $\hat{R}_T$ تعدّل $\hat{R}$ بتقدير عيّنة للتصحيح الذي نحتاجه لمعالجة الانحياز، ويتطابق $\hat{R}_B$ و $\hat{R}_T$ إلى حدود من مرتبة $\frac{1}{n}$ وإنجازاهما متشابهان جدًّا.

جدول (٦-٦) مجتمع اصطناعي صغير

	الطبقة					
	I		II		III	
	y	x	y	x	y	x
	2	2	2	1	3	1
	3	4	5	4	7	3
	4	6	9	8	9	4
	11	20	24	23	25	12
المجاميع	20	32	40	36	44	20
R_h	0.625		1.111		2.200	

(٦-١٦) مقارنـة الطـرق

مثـال

يتضمن المجتمع الاصطناعي في الجدول (٦-٦) ثلاث طبقات حيث $N_h=4$ و $n_h=2$ في كل طبقة. وقد وُضع المجتمع عن عَمد بحيث (ا) يتغير $\bar{R}_h$ بصورة ملحوظة من طبقة إلى طبقة وبالتالي يجعل التقدير النسبة المنفصل $\hat{Y}_{Rs}$ التقدير المفضلّ و(ب) يبالغ $\hat{R}_h$ في التقدير في كل طبقة مع خطر تجمّع انحياز جدّي في $\hat{Y}_{Rs}$. وقد قورنت الطرق التالية لتقدير مجموع المجتمع Y .

نشر بسيط : $\sum N_h \bar{y}_h$

نسبة مركبة : $(\bar{y}_{st}/\bar{x}_{st})X$

نسبة منفصلة : $\sum (\bar{y}_h/\bar{x}_x)X_h = \sum \hat{R}_h X_h$

والتقديرات الباقية، Hartley-Ross المنفصل، Lahiri ، وطرق Beale, Quenouille و Tin يتخذ فيها التقدير النسبة المنفصل الشكل نفسه، باستثناء أن $\hat{R}_{(HR)h}$ ، $\hat{R}_{(L)h}$ وما شابه، توضع بدلاً من $\hat{R}_h$. (من أجل $n_h=2$ تكون طريقتا Hartley-Ross و Mickey متطابقتين). ويوجد $6^3=216$ من العيّنات الممكنة. والنتائج مضبوطة في حدود أخطاء تدوير الأرقام العشرية. وأنا مدين للدكتور Joseph Sedransk لمساعدته في بعض الحسابات.

جدول (٦-٧) نتائج تقديرات مختلفة لـ Y

الطريقة	التباين	مربع الانحياز	متوسط مربعات الخطأ
نشر بسيط	820.3	0.0	820.3
نسبة مركبة	262.8	6.5	269.3
نسبة منفصلة	35.9	24.1	60.0
هارتلي - روس منفصل	153.6	0.0	153.6
لاهيري - منفصل	19.6	0.0	19.6
كوينويل - منفصل	42.9	1.1	44.0
بيل - منفصل	28.9	8.0	36.9
تين - منفصل	28.6	5.7	34.3

وتبيّن النتائج في الجدول (٦-٧) بعض النواحي المثيرة للاهتمام. ففي التقدير النسبة المركب نجد أن مساهمة مربع الانحياز في متوسط مربعات الخطأ هي مساهمة تافهة، وذلك بالرغم من الشروط المتطرفة، إلا أن هذا التقدير يتصف بالضعف فيما يتعلق بالتباين وذلك بسبب التغير الواسع في المقادير $\hat{R}_h$. وكما يقضي متوسط مربعات الخطأ نجد أن التقدير النسبة المنفصل أدق بكثير من التقدير المركب، إلا أن انحيازه رديء. ومن بين الطرق غير المنحازة تُظهر طريقة Hartley-Ross تبايناً مرتفعاً نسبياً، حيث وُجد أنها تفعل ذلك عندما يكون $n_h=2$ في بعض الدراسات المتعلقة بمجتمعات فعلية. وتؤدي طريقة Lahiri أداءً حسناً بصورة خاصة. ويناسب هذا المجتمع

طريقة Lahiri بسبب أن إحدى الوحدات في كل طبقة تتضمن قيمًا مرتفعة بصورة غير عادية لكل من x_i و y_i مما يؤدي إلى احتمال مرتفع لسحب هذه الوحدة، وتعطي العيّنات التي تتضمن هذه الوحدة تقديرات جيدة لـ $\hat{R}_h$.

وقد أدّت طرق Beale, Quenouille و Tin جميعها إلى تخفيضات كبيرة في الانحياز بالمقارنة مع التقدير النسبة المنفصل، وكان لها جميعًا متوسطات خطأ تربيعي أصغر، وهكذا تكون قد أنجزت أهدافها الرئيسة في هذا المثال.

ودراسة J.N.K. Rao (1969) لمجتمعات واقعية، وهي الدراسة المذكورة في الفقرة (٦-٩) قارنت بين طرق Beale, Quenouille و Tin من أجل $n=2,4,6,8$ وذلك في 15 من هذه المجتمعات. وفي حالة $n=2$ كان الاختبار الأكثر قسوة، إذ كان الوسيط والربيع الأعلى للمقدار $\frac{\text{الانحياز}}{\sqrt{\text{متوسط مربعات الخطأ}}}$ كما يلي: $\hat{R}_Q$، 3%، 7% ؛ $\hat{R}_B$ ؛ 8%، 12% ؛ $\hat{R}_T$ ؛ 8%، 19% وذلك في مقابل 15% و 20% لـ $\hat{R}$. وفيما يتعلق بالانحياز تُقدم الطرق الأكثر تعقيدًا عونًا جديرًا بالذكر في مثل هذه العيّنات الصغيرة.

وقد قارنت الدراسة نفسها متوسطات الخطأ التربيعي لخمس من التقديرات في هذه الفقرة مع متوسط الخطأ التربيعي لـ $\hat{R}$. (حُذفت طريقة Lahiri باعتبار أن الدراسة تقتصر على عيّنات عشوائية بسيطة). ولكل طريقة حُسبت النسبة $100\ MSE(\hat{R}_Q)/MSE(\hat{R})$* في كل مجتمع.

وفي حالة $n=4$ كانت تقديرات Quenouille و Mickey في هذه المجتمعات، متخلفة قليلاً عن $\hat{R}$ إلا أنه عندما $n\geqslant 6$ كان متوسط متوسطات الأخطاء التربيعية لكل طريقة قريبًا جدًّا من نظيره الخاص بـ $\hat{R}$. وفي عيّنة صغيرة جدًّا من مجتمع بمفرده تقترح الدراسة أن لا فائدة لمثل هذه الطرق المعقدة من حيث توفيرها لدقة تفوق بصورة ملموسة دقة $\hat{R}$. ولكن حقيقة أنها تُخفض الانحياز في طبقة بمفردها دون زيادة متوسط الخطأ التربيعي أو بزيادة بسيطة فيه ينبغي أن تمنحها ميزة في حالة التقدير النسبة المنفصل مع عدد كبير من الطبقات وعيّنة صغيرة من كل طبقة.

---

* ترمز MSE لمتوسط الخطأ التربيعي. المترجم

وتحت نموذج انحدار خطي قام P.S.R.S. Rao و J.N.K. Rao (1971) ؛ Hutchinson (1971) ؛ Kuzik (1974) بمقارنات بين متوسطات مربعات الخطأ لهذه الطرق في حالة n صغير، وقد أعطت هذه المقارنات نتائج تتفق بصورة عامة مع تلك التي حصلنا عليها من مجتمعات فعلية.

(٦-١٧) تقدير محسّن للتباين

اقترح Tukey (1958) تقديرًا يستحق الاعتبار لتباين تقدير مدية الجيب $\hat{R}_Q$ (تقدير Quenouille) في حالة عيّنات صغيرة أو معتدلة الحجم. فمع g من الفئات، $n=mg$ ، وf قابل للإهمال، يكون $\hat{R}_Q$ هو متوسط المقادير، وعددها g ،

$$\hat{R}'_j = g\hat{R} - (g-1)\hat{R}_f$$

حيث نحسب $\hat{R}_j = \sum y / \sum x$ بعد حذف الفئة j . وإذا أمكن اعتبار المقادير $\hat{R}'_j$، وعددها g ، تقديرات مستقلة لِـ R فعندئذ، ومع معاينة عشوائية بسيطة، يمكن أخذ

$$v(\hat{R}_Q) = \frac{(1-f)}{g} \frac{\sum^{g} (\hat{R}'_j - \hat{R}_Q)^2}{(g-1)} \tag{6.85}$$

تقديرًا غير منحاز لِـ $V(\hat{R}_Q)$.

وبما أن $\hat{R}'_j - \hat{R}_Q = -(g-1)(\hat{R}_j - \hat{R}_-)$, فيمكن حساب (6.85) ، بسهولة أكبر، من

$$v(\hat{R}_Q) = \frac{(1-f)(g-1)}{g} \sum^{g} (\hat{R}_j - \hat{R}_-)^2 \tag{6.86}$$

حيث $\hat{R}_-$ هو متوسط المقادير $\hat{R}_j$ ، وعددها g .

والمقادير $\hat{R}$ أو $\hat{R}'_j$ مع قيم مختلفة لِـ j ، هي بالطبع غير مستقلة، والعلاقة (6.86) هي علاقة تقريبية. وقد أرسينا حتى الآن الخواص التحليلية لِـ $v(\hat{R}_Q)$ في عيّنات كبيرة فقط. وبيّن Arvesen (1969) أنه في صف واسع من التقديرات المتناظرة في عناصر العيّنة، وهو صف يتضمن $\hat{R}_Q$ ، [تقديرات تسمى إحصاءات-U

لِـ Hoeffding (1948)] تصبح العلاقة $v(\hat{R}_Q)$ غير منحازة إما من أجل g مثبّت أو من أجل $g=n$ عندما يصبح n كبيراً.

ومن دراسة Rao (1969) لثمانية مجتمعات فعلية ذكرنا في الفقرة (٦-٩) متوسط النسبة المئوية للتقدير بالنقصان في $v_1(\hat{R})$ كتقدير للقيمة الصحيحة لِـ $MSE(\hat{Y})$ وذلك في حالة $n=4,6,8,12$ وللمقارنة يبيّن الجدول (٦-٨) المتوسطات المقابلة للنسب المئوية للانحيازات في $v(\hat{R}_Q)$ وتمثل هذه متوسطات الأعداد الثمانية $100[v(\hat{R}_Q)-MSE(\hat{R}_Q)]/MSE(\hat{R}_Q)$ محسوبة من المجتمعات الثمانية.

جدول (٦ ـ ٨) متوسط النسبة المئوية للانحياز في تقديرات تباين.

متوسط المقدار	$n=$ 4	6	8	12
$100[v(\hat{R}_Q)-MSE(\hat{R}_Q)]/MSE(\hat{R}_Q)$	+11%	+10%	+6%	+1%
$100[v(\hat{R}_Q)-MSE(\hat{R})]/MSE(\hat{R})$	+11%	+10%	+6%	+1%
$100[v_1(\hat{R})-MSE(\hat{R})]/MSE(\hat{R})$	−31%	−23%	−21%	−18%

ويعطي الجدول (٦-٨) أيضًا متوسطات المقادير $100\ [v(\hat{R}_Q)-MSE(\hat{R})]/MSE(\hat{R})$ وهذه المتوسطات ذات فائدة بالنسبة لباحث يستخدم $\hat{R}$ إلا أنه يرحب بوضع $v_1(\hat{R})$ بدلاً من $v(\hat{R}_Q)$ كتقدير لِـ $V(\hat{R})$ إذا بدا له أنه أقل انحيازًا. ونظرًا لانحياز كل من $\hat{R}$ و $\hat{R}_Q$ تمت مقارنة $V(\hat{R}_Q)$ من خلال متوسطات مربعات الخطأ باعتبارها مناسبة أكثر.

في هذه المجتمعات يبالغ $v(\hat{R}_Q)$ بصورة طفيفة في تقدير كل من $MSE(\hat{R}_Q)$ و $MSE(\hat{R})$، بينما ينحاز $v_1(\hat{R})$ انحيازات سالبة كبيرة في هذه العيّنات الصغيرة.

ومتخذين مربعات معاملات اختلاف تقديرات التباين أداة للحكم، يمكن القول إن استقرار $v(\hat{R}_Q)$ كان ضعيفًا في هذه العيّنات بالمقارنة مع استقرار $v_1(\hat{R})$, .

وعلى أي حال، ففي دراسات Rao و Beegle (1968) لِـ $v(\hat{R}_Q)$ و $v_1(\hat{R})$ تحت نموذج انحدار خطي لِـ y على x في مجتمع لانهائي، وحيث يتوزع x وفق التوزيع الطبيعي،

تبيّن أن $v(\hat{R}_Q)$ و $v_1(\hat{R})$ يتمتعان بالاستقرار نفسه تقريبًا عندما تتراوح قيمة n بين 4 و 12 .

وفي حالة التقدير النسبة المنفصل وطبقات متعددة تقترح هذه النتائج أن $\sum X_h^2 v(\hat{R}_{Oh})$ متفوق على $\sum X_h^2 v_1(\hat{R}_h)$ كتقدير لِـ $V(\hat{Y}_{Rs})$. ومن المحتمل أن يكون $\sum X_h^2 v(\hat{R}_{Oh})$ أقل انحيازًا . وينبغي أن يكون لكليهما استقرار مناسب. إلا أنه في حالة عدد قليل من الطبقات فقط تبقى المسألة موضع تساؤل حتى ظهور المزيد من المقارنات .

(٦ ـ ١٨) مقارنة نسبتين

كثيرًا ما يكون من الضروري في المسوح التحليلية تقدير الفرق $\hat{R}-\hat{R}'$ بين نسبتين، وحساب الخطأ المعياري لِـ $\hat{R}-\hat{R}'$. والعلاقات المعطاة هنا هي علاقات خاصة بتقدير تباين $\hat{R}-\hat{R}'$ ، باعتبارها العلاقات المطلوبة بصورة عامة. وقد حذفت حدود الـ ت م م لأسباب قُدمت في الفقرة (٢-١٤) .

ونفترض أولاً عيّنة عشوائية بسيطة. ويمكن تمييز ثلاث حالات .

النسبتان مستقلتان

ويقع هذا عندما تكون الوحدات مصنفة إلى صنفين متميزين ونرغب في مقارنة نسب قدرناها بصورة منفصلة في الطبقتين. وعلى سبيل المثال، في دراسات مصاريف الأسرة، يمكن تقسيم عيّنة عشوائية بسيطة إلى منازل مملوكة ومُستأجرة كي نقارن الدخل المصروف على صيانة المنزل في الصنفين. وإذا رمزنا للنسب المقدّرة بِـ $\hat{R}=\bar{y}/\bar{x}$ ، $\hat{R}'=\bar{y}'/\bar{x}'$ فعندئذ :

$$v(\hat{R}-\hat{R})=v(\hat{R})+v(\hat{R}') \tag{6.87}$$

للنسبتين المقام نفسه

وعندما تكون الوحدة عنقودًا من الأسر، فقد نرغب في مقارنة نسبة الذكور البالغين الذين يستخدمون آلة حلاقة كهربائية مع نسبة الذين يستخدمون شفرات

الحلاقة. وفي أي وحدة، y = عدد الذكور البالغين الذين يستخدمون آلة حلاقة كهربائية، و y'= عدد الذكور البالغين الذين يستخدمون شفرات حلاقة، و x= العدد الكلي للذكور البالغين.

$$\hat{R}-\hat{R}'=\frac{\bar{y}-\bar{y}'}{\bar{x}}$$

وإذا كان $d_i=y_i-y_i'$ فيمكن حساب تقدير تباين $\hat{R}-\hat{R}'$ على الشكل،

$$v(\hat{R}-\hat{R}')\doteq\frac{1}{n(n-1)\bar{x}^2}\sum_{i=1}^{n}[d_i-(\hat{R}-\hat{R}')x_i]^2 \qquad (6.88)$$

للنسبتين مقامان مختلفان إلا أنهما قد تكونان مرتبطتين

كمثال نذكر مقارنة نسبة الذكور المدخنين مع نسبة النساء المدخنات، في مسح تكون فيه الوحدة عنقودًا من المنازل. ومن الناحية الرياضية هذه هي الحالة الأعم.

$$v(\hat{R}-\hat{R}')=v(\hat{R})+v(\hat{R}')-2\operatorname{cov}(\hat{R}\hat{R}') \qquad (6.89)$$

والحد الوحيد غير المألوف هو $\operatorname{cov}(\hat{R}\hat{R}')$ لنكتب حسب الطريقة المعتادة،

$$\hat{R}-R\doteq\frac{\bar{y}-R\bar{x}}{\bar{X}} \qquad \hat{R}'-R'\doteq\frac{\bar{y}'-R'\bar{x}'}{\bar{X}'}$$

فنجد،

$$\operatorname{cov}(\hat{R}\hat{R}')\doteq\frac{1}{n\bar{X}\bar{X}'}\operatorname{cov}(y_i-Rx_i)(y_i'-R'x_i')$$

ويمكن حساب تقدير عيّنة كما يلي:

$$\operatorname{cov}(\hat{R}\hat{R}')\doteq\frac{1}{n(n-1)\bar{x}\bar{x}'}\sum^{n}(y_iy_i'-\hat{R}y_i'x_i-\hat{R}'y_ix_i'+\hat{R}\hat{R}'x_ix_i') \qquad (6.90)$$

مثال: نُفذت التجارب الحقلية عام 1954 من أجل لقاح Salk لشلل الأطفال بين أطفال الصفوف الثلاثة الأولى في جميع المدارس في عدد من المناطق. ولم يجر اختيار المناطق عشوائيًا، إذ فُضلت المناطق التي لها تاريخ هجمات سابقة لمرض شلل الأطفال، إلا أننا سنفترض في هذا المثال أنها عيّنة عشوائية من مجتمع ما.

والأطفال الذين لم يسمح آباؤهم باشتراكهم في التجربة سيُدعون «غير الملقحين» وبـالطبع، لم يتلقوا زُرقات. وقد أعطي نصف الأطفال الذين سُمح لهم بالاشتراك ثلاث زرقات من سائل معقم. ودُعيت هذه الفئة المحايدة (Placebo) ومن المعلومات الإحصائية في الجدول (٦-٩) سنقارن التكرارين $\hat{R}$ و $\hat{R}'$ لحالات شلل الأطفال في فئة غير الملقحين والفئة المحايدة. ولتخفيف حجم البيان الإحصائي ستقتصر المقارنة على 43 منطقة في كل منها أكثر من 4000 طفل في الفئتين معًا.

جدول (٦-٩) عدد الأطفال (x,x') وحالات الشلل (y,y') للمنطقة الواحدة

x*	x'	y†	y'		x	x'	y	y'
4.1	2.4	0	0		13.8	25.6	3	3
3.5	8.0	1	6		10.5	8.1	2	0
4.1	6.1	7	2		21.6	25.9	10	7
2.6	4.6	2	1		3.5	6.7	2	2
2.4	1.5	2	1		6.8	7.3	3	8
2.2	1.9	0	0		2.3	3.7	0	1
1.1	4.0	1	1		2.6	2.9	2	0
1.6	4.0	1	2		6.0	11.1	3	1
5.7	7.8	1	4		11.0	14.8	7	11
3.3	11.0	3	7		19.4	42.5	11	14
1.0	3.8	0	1		6.8	13.7	6	2
2.0	5.2	1	0		1.2	4.0	3	1
8.3	19.0	4	4		5.4	9.3	11	6
1.0	3.7	1	5		1.7	2.6	0	2
1.1	4.2	0	1		2.1	2.3	0	0
2.3	6.8	1	2		1.5	2.6	0	0
1.9	3.5	0	2		3.0	4.0	0	2
				Totals	167.4	284.6	88	99

* $x=x,x'$ عدد الأطفال في الفئة المحايدة و x' عدد الأطفـال في فئـة غير الملحقـين (بـالآلاف).

† $y=y,y'$ عدد حالات الإصـابـة بالشلل في الفئـة المحـايدة، و y' عدد الإصـابـات في فئـة (غـير الملحقين).

وفي هذا البيان سيُنتج أي تغير في معدل الإصابة بالشلل من منطقة إلى منطقة ارتباطًا موجبًا بين $\hat{R}$ و $\hat{R}'$.

والمقادير التالية مستمدّة من المجاميع .

(Placebo) المحايدة $\bar{x}=\frac{167.4}{34}=4.9235 \qquad \hat{R}=\frac{88}{167.4}=0.525687,$

غير ملقّح $\bar{x}'=\frac{284.6}{34}=8.3706 \qquad \hat{R}'=\frac{99}{284.6}=0.347857,$

ومن أجـل $v(\hat{R})$ ، $v(\hat{R}')$ ، و$\text{cov}\,(\hat{R}\hat{R}')$ نحتـاج إلى جميع مجاميع المـربعات والجداءات غير المصححة بين المتغيرات الأربعة .

$$v(\hat{R})=\frac{1}{n(n-1)\bar{x}^2}\left(\sum y^2-2\hat{R}\sum yx+\hat{R}^2\sum x^2\right)$$

$$=\frac{1}{(34)(33)(4.9235)^2}[(564)-(1.05137)(822.2)+(0.27635)(1661.92)]$$

$$=0.00584$$

وبصورة مماثلة نجد $v(\hat{R}')=0.00240$

$$\text{cov}\,(\hat{R}\hat{R}')=\frac{1}{n(n-1)\bar{x}\bar{x}'}\left(\sum yy'-\hat{R}\sum y'x-\hat{R}'\sum yx'+\hat{R}\hat{R}'\sum xx'\right)$$

$$=\frac{(497)-(0.52569)(844.6)-(0.34786)(1397.4)+(0.52569)(0.34786)(2690.8)}{(34)(33)(4.9235)(8.3706)}$$

$$=0.00127$$

ومنه

$$\text{s.e.}(\hat{R}-\hat{R}')=\sqrt{0.00584+0.00240-0.00254}=0.0754$$

وبـما أن $\hat{R}-\hat{R}'=0.1778$ فالفـرق يقـترب من أن يكون مهمًا عند مستوى الأهمية 5% (قـد يكـون توزيع $\hat{R}-\hat{R}'$ منحـازًا، إلى حد ما مع مثل هذا الحجم للعيّنة) . والتفسير الممكن هو أنه قد يكون للأطفال غير الملقحين مناعة طبيعية أكبر ضد الشلل من أطفال الفئة المحايدة (Placebo) .

وقد تنشأ المسألة نفسها في عيّنات طبقية تتقاطع فيها ميادين الدراسة مع جميع الطبقات . وإذا بدا أن $\hat{R}_h$ ، $\hat{R}_h'$ تتغير من طبقة إلى طبقة فمن المحتمل أن المقارنة

ستبنى على دراسة قيم $\hat{R}_h - \hat{R}_h'$ في كل طبقة بمفردها، وبإيجاد الأخطاء المعيارية لـ $\hat{R}_h - \hat{R}_h'$ يمكن تحديد ما إذا كانت هذه الفروق تتغير من طبقة إلى طبقة، وإذا لم يكن الأمر كذلك نحسب فرقًا إجماليًا يمكن الاطمئنان إليه.

وإذا لم تُظهر $\hat{R}_h'$ و $\hat{R}_h$ تغيرًا حقيقيًا من طبقة إلى أخرى فقد يكون حساب التقديرات المركبة $\hat{R}_c$ و $\hat{R}_c'$ كافيًا وكما رأينا من قبل:

$$v(\hat{R}_c - \hat{R}_c') = v(\hat{R}_c) + v(\hat{R}_c') - 2\,\text{cov}\,(\hat{R}_c\hat{R}_c') \tag{6.91}$$

وبوضع $d_{hi} = (y_{hi} - \bar{y}_h) - \hat{R}_c(x_{hi} - \bar{x}_h)$, نجد

$$v(\hat{R}_c) = \frac{1}{\bar{x}_{st}^2} \sum_h \frac{N_h^2}{n_h(n_h - 1)} \sum_i d_{hi}^2 \tag{6.92}$$

$$\text{cov}\,(\hat{R}_c\hat{R}_c') = \frac{1}{\bar{x}_{st}\bar{x}_{st}'} \sum_h \frac{N_h^2}{n_h(n_h - 1)} \sum_i d_{hi}d_{hi}' \tag{6.93}$$

ويقدّم Kish و Hess (1959b) مناقشة أكثر كمالًا لمقارنة نسب، وتتضمن هذه المناقشة صيغًا حسابية مختزلة وذلك عندما تسمح العيّنة بمثل هذه الصيغ.

(٦-١٩) نسبة نسب

نريد في بعض التطبيقات تقدير النسبة R/R' لنسبتين. وهكذا فقد نهتم في المثال السابق (فقرة ٦-١٨)، بالنسبة R/R' لمعدلي حالات الإصابة بالشلل بين أطفال الفئة المحايدة وفئة غير الملقحين. أو نسبة نسبتي الذكور والإناث في القوة العاملة مستخدمين عيّنة عنقودية. وإذا كان البيان الإحصائي حول (y,x) متوافرًا من العيّنة نفسها في فترتين زمنيتين مختلفتين فقد تكون النسبة هي نسبة المصاريف الأسبوعية على الطعام للأسرة الواحدة في الفترتين الزمنيتين.

ومع عيّنة عشوائية بسيطة (مثلًا، عيّنة من العناقيد) يكون تقدير العيّنة لـ R/R' هو $\hat{R}/\hat{R}' = (\bar{y}/\bar{x})/(\bar{y}'/\bar{x}')$ ويدعى أحيانًا التقدير النسبة المضاعف. وكما في حالة نسبة بمفردها نجد أن الحد الرئيس من انحياز $\hat{R}/\hat{R}'$ هو من مرتبة $\frac{1}{n}$ إلا أنه أكثر تعقيدًا مما هو في حالة $\hat{R}'$ أو $\hat{R}$ ويمكننا كتابة

$$\hat{R}/\hat{R}' = (R/R')(1+\delta\bar{y})(1+\delta\bar{x}')/(1+\delta\bar{x})(1+\delta\bar{y}') \qquad (6.94)$$

حيث يرمز $\delta\bar{y}$ لـ $(\bar{y}-\bar{Y})/\bar{Y}$ وهلمجرا. وعند نشرهذهالعبارة تدخل ستة حدود تربيعية من مرتبة $\frac{1}{n}$ في انحياز $\hat{R}/\hat{R}'$. ويعطي Rao و Pereira (1968) عبارة مضبوطة للانحياز.

ويمكن كتابة العلاقة (6.8) الخاصة بالتباين النسبي $V(\hat{R})/R^2 = C_{\hat{R}\hat{R}}$ لنسبة كما يلي:

$$C_{\hat{R}\hat{R}} = C_{\bar{y}\bar{y}} + C_{\bar{x}\bar{x}} + 2C_{\bar{y}\bar{x}}$$

ومنه نجد أن الحد الرئيس في $V(\hat{R}/\hat{R}')$ هو،

$$V(\hat{R}/\hat{R}') \doteq \left(\frac{R}{R'}\right)^2 (C_{\hat{R}\hat{R}} + C_{\hat{R}'\hat{R}'} - 2C_{\hat{R}\hat{R}'}) \qquad (6.95)$$

حيث

$$C_{\hat{R}\hat{R}'} = C_{\bar{y}\bar{y}'} + C_{\bar{x}\bar{x}'} - C_{\bar{y}\bar{x}'} - C_{\bar{y}'\bar{x}} \qquad (6.96)$$

ولإيجاد تقدير العيّنة الموافق $v(\hat{R}/\hat{R}')$ نعوّض تقديرات العيّنة للحدود المذكورة في (6.95).

مثال

في نسبة حالات الشلل بين الفئة المحايدة و غير الملقحين نجد $\hat{R}/\hat{R}' = 0.52569/0.34786 = 1.511$. ولتقدير الخطأ المعياري لهذه النسبة نجد من الحسابات في المثال السابق:

$$\hat{C}_{\hat{R}\hat{R}} = \frac{0.00584}{(0.5257)^2} = 0.0211; \qquad \hat{C}_{\hat{R}'\hat{R}'} = \frac{0.00240}{(0.3479)^2} = 0.0198;$$

$$\hat{C}_{\hat{R}\hat{R}'} = \frac{0.00127}{(0.5257)(0.3479)} = 0.00694$$

$$v(\hat{R}/\hat{R}') = (1.511)^2(0.0211 + 0.0198) - 0.0139) = 0.0617$$

$$\text{s.e.}\ (\hat{R}/\hat{R}') = 0.248$$

ومن حين لآخر يُستخدم التقدير النسبة المضاعف بدلاً من $\hat{Y}_R = \hat{R}X$ لتقدير مجموع مجتمع Y وذلك وفقًا لما اقترحه Keyfitz (1960, yates) فلنفترض أن $\hat{R}' = (\bar{y}'/\bar{x}')$ معروفة من العيّنة نفسها المأخوذة في فترة سابقة، وأن $R' = \bar{Y}'/\bar{X}'$ معروف أيضًا. وإذا وجدنا، مثلاً، أن $\hat{R}' / R'$ أكبر بقليل من الواحد. فيمكن اللجوء إلى المحاكمة البدهية التالية: فنقول إنه من المرجّح أيضًا أن يعطي $\hat{R}$ تقديرًا بالزيادة لـ R وإنه ينبغي تعديله في اتجاه التخفيض بقسمته على النسبة $\hat{R}' / R'$. ويقود هذا إلى التقدير النسبة المضاعف $\hat{Y}_{DR}$.

$$\hat{Y}_{DR} = \frac{R'}{\hat{R}'}(\hat{R}X) = \frac{\hat{R}}{\hat{R}'}(R'X) \tag{6.97}$$

وبما أن التباين النسبي لـ $\hat{Y}_R$ هو C_{RR} بينما التباين النسبي لـ $\hat{Y}_{DR}$ هو،

$$C_{RR} + C_{R'R'} - 2C_{RR'} \tag{6.98}$$

فستعطي النسبة المضاعفة تقديرًا أكثر دقة في العيّنات الكبيرة إذا كان الارتباط بين $\hat{R}$ و R' مرتفعًا بصورة كافية.

(٦-٢٠) التقديرات النسبة لعدة متغيرات

عمّم Olkin (1958) تقدير النسبة، إلى الحالة التي يتوافر فيها p من المتغيرات المساعدة $(x_1, x_2, \ldots, x_p)$ وفي حالة مجموع المجتمع يكون التقدير المقترح، ولنرمزله بـ $\hat{Y}_{MR}$ (Multivariate Ratio) .

$$\begin{aligned}\hat{Y}_{MR} &= W_1 \frac{\bar{y}}{\bar{x}_1} X_1 + W_2 \frac{\bar{y}}{\bar{x}_2} X_2 + \cdots + w_p \frac{\bar{y}}{\bar{x}_p} X_p \\ &= W_1 \hat{Y}_{R_1} + W_2 \hat{Y}_{R_2} + \cdots + W_p \hat{Y}_{R_p}\end{aligned}$$

حيث الـ W_i ترجيحات سنحدد قيمتها بحيث تكون دقة $\hat{Y}_{MR}$، أعظم ما يمكن، خاضعة للشرط $\sum W_i = 1$. ويبدو هذا النوع من التقدير مناسبًا عندما يكون انحدار y على $(x_1, x_2, \ldots, x_p)$ خطيًا ويمرّ من المبدأ. ويجب أن تكون مجاميع المجتمع X_i معروفة.

وسنستعرض الطريقة في حالة متغيرين مساعدين (x) باعتبار أنها ينبغي أن تكون الحالة الأكثر تواترًا في التطبيقات.

لدينا

$$\hat{Y}_{MR} - Y = W_1(\hat{Y}_{R_1} - Y) + W_2(\hat{Y}_{R_2} - Y)$$

وبالتالي، مفترضين انحيازًا قابلاً للإهمال نجد

$$\begin{aligned} V(\hat{Y}_{MR}) &= W_1^2 V(\hat{Y}_{R_1}) + 2W_1 W_2 \operatorname{cov}(\hat{Y}_{R_1}\hat{Y}_{R_2}) + W_2^2 V(\hat{Y}_{R_2}) \\ &= W_1^2 V_{11} + 2W_1 W_2 V_{12} + W_2^2 V_{22} \end{aligned} \quad (6.99)$$

حيث $V_{11} = V(\hat{Y}_{R_1})$ الخ . وتحت الشرط $W_1+W_2=1$ نجد أن القيم w_2, w_1 التي تجعل التباين أصغر ما يمكن هي :

$$W_1 = \frac{V_{22} - V_{12}}{V_{11} + V_{22} - 2V_{12}}, \qquad W_2 = \frac{V_{11} - V_{12}}{V_{11} + V_{22} - 2V_{12}}$$

والتباين الأصغري هو،

$$V_{min}(\hat{Y}_{MR}) = \frac{V_{11}V_{22} - V_{12}^2}{V_{11} + V_{22} - 2V_{12}} \quad (6.100)$$

وفي حالة ρ من المتغيرات ننظر إلى حساب V^{ij} عكس المصفوفة V_{ij} وعندئذ يكون الحل الأمثل $W_i = \sum_i / \sum$ حيث $\sum_i$ مجموع العناصر في العمود i من المصفوفة V^{ij} و $\sum$ مجموع كافة الـ p^2 من العناصر في المصفوفة V^{ij} ويكون التباين الأصغري $1/\sum$.

وفي التطبيقات نحدّد الترجيحات باستخدام تقديرات v_{ij} للتباينات والتغايرات .
ومن (6.7) في الفقرة (٦-٣) نجد،

$$v_{11} = \frac{(1-f)\hat{Y}^2}{n}(c_{yy} + c_{11} - 2c_{y1})$$

$$v_{22} = \frac{(1-f)\hat{Y}^2}{n}(c_{yy} + c_{22} - 2c_{y2})$$

حيث $c_{yy} = s_y^2/\bar{y}^2$ ويمكن التعبير عن التغاير على الشكل،

$$v_{12} = \frac{(1-f)\hat{Y}^2}{n}(c_{yy} + c_{12} - c_{y1} - c_{y2})$$

وإحدى الطرق المريحة للحسابات هي الحصول أولاً على المصفوفة،

$$C=\begin{pmatrix} c_{yy} & c_{y1} & c_{y2} \\ c_{y1} & c_{11} & c_{12} \\ c_{y2} & c_{12} & c_{22} \end{pmatrix}$$

وإذا كان $v_{ij}'=nv_{ij}/(1-f)\hat{Y}^2$ ، فيمكن الحصول بسهولة على المصفوفة v'_{ij} بأخذ مقارنات قطرية في C ، أي

$$v_{11}'=c_{yy}+c_{11}-c_{y1}-c_{y1}$$

$$v_{12}'=c_{yy}+c_{12}-c_{y1}-c_{y2}$$ إلخ .

ولا نحتاج إلى العامل $(1-f)\ \hat{Y}^2/n$ عند حساب المقادير w_i ، إلا أنه يجب إدخاله عند حساب التباين الأصغري . وهكذا نجد،

$$v_{min}(\hat{Y}_{MR})=\frac{(1-f)\hat{Y}^2}{n}\frac{(v_{11}'v_{22}'-v_{12}'^2)}{(v_{11}'+v_{22}'-2v_{12}')} \tag{6.101}$$

وبالنظر إلى مقدار الحسابات التي تتضمنها هذه الطريقة فمن المرجّح أن استخدام هذا التقدير سيقتصر على أصغر المسوح الإحصائية ذات الهدف المتخصص . والطريقة مؤهلة لأن تعطي زيادة ملحوظة في الدقة فوق $\hat{Y}_{R_1}$ أو $\hat{Y}_{R_2}$ كل بمفرده .

(٦-٢١) المقدّرات الجدائية

إذا كان x متغيراً مساعدًا يرتبط سلبًا بـ y حيث x و y متغيران يأخذان فقط قيمًا موجبة ، فالشّبَه الطبيعي للمقدّر النسبة هو المقدّر الجدائي الذي نأخذ فيه،

$$\hat{\bar{Y}}_p=\frac{\bar{x}}{\bar{X}}\bar{y};\qquad \hat{Y}_p=N\frac{\bar{x}}{\bar{X}}\bar{y} \tag{6.102}$$

وباستخدام النشر العادي وفقًا لسلسلة تايلور نجد أن العلاقة المشابهة لِـ (6.8) في حالة المقدّر الجدائي من عيّنة عشوائية بسيطة كبيرة هي ،

$$(cv)^2=\frac{(1-f)}{n}(C_{yy}+C_{xx}+2C_{yx}) \tag{6.103}$$

حيث $(cv)^2$ هو مربع معامل الاختلاف لأي من $\hat{Y}_p$ أو $\hat{\bar{Y}}_p$ وقد وسّع P.S.R.S. Rao و Mudholkar (1967) المقـدّر النسبة متعدد المتغيرات لِـ Olkin إلى مركّب مرجّح من

المقدّرات النسبة (في حالة x_i مرتبطة إيجابًا بـ y) ومقدّرات جدائية (في حالة x_i مرتبط سلبًا بـ y).

تماريـن

(٦-١) أعطى مسح استكشافي لـ 21 أسرة المعلومات التالية عن أعداد الأفراد (x)، الأطفال (y_1)، العربات (y_2) وأجهزة التلفاز (y_3).

x	y_1	y_2	y_3	x	y_1	y_2	y_3	x	y_1	y_2	y_3
5	3	1	3	2	0	0	1	6	3	2	0
2	0	1	1	3	1	1	1	4	2	1	1
4	1	2	0	2	0	2	0	4	2	1	1
4	2	1	1	6	4	2	1	3	1	0	1
6	4	1	1	3	1	0	0	2	0	2	1
3	1	1	2	4	2	1	1	4	2	1	1
5	3	1	1	5	3	1	1	3	1	1	1

مفترضًا أن مجموع المجتمع X معروف، هل توصى باستخدام التقديرات النسبة بدلًا من النشر البسيط لتقدير الأعداد الكلية للأطفال، العربات، وأجهزة التلفاز؟

(٦-٢) في حقل شعير، قمنا بوزن الحبوب الناتجة y_i والحبوب بالإضافة إلى القشّ x_i وذلك لكل من عدد كبير من وحدات المعاينة التي حُدّد موضعها عشوائيًا فوق الحقل. وقمنـا أيضًـا بوزن الإنتـاج الكلي (حبوب + قشّ) لمجمل الحقل. وقد حصلنا على المعلومات التالية:

$$c_{xx} = 1.11, \quad c_{yx} = 0.78, \quad c_{yy} = 1.13$$

احسب الكسب الحاصل في الدقة نتيجة لتقدير إنتاج الحبوب من نسبة الحبوب إلى الإنتاج الكلي بدلًا من متوسط إنتاج الحبوب في الوحدة.

ونحتاج إلى 20 دقيقة حتى ندرس الحبوب ونزنها في كل وحدة، دقيقتين لوزن القش في كل وحدة، وساعتين لجمع ووزن الإنتاج الكلي للحقل. فكم من الوحدات لكل حقل يجب أخذها حتى يكون التقدير النسبة أكثر اقتصادًا من متوسط إنتاج الوحدة؟

(٦-٣) في العيّنة الإحصائية في الجدول (٦-١)، لدينا $\hat{Y}_R = 28,367$ و $c_{yy} = 0.0142068$ و $c_{yz} = 0.0146541$ ، $c_{zz} = 0.0156830$

احسب 95% حدود ثقة تربيعية لِـ Y وقارنها مع الحدود الناتجة عن التقريب الطبيعي.

(٦-٤) قيست قيم y و x لكل وحدة من وحدات عيّنة عشوائية بسيطة من مجتمع. إذا كان $\bar{x}$ متوسط المجتمع x معروفًا فأي الطرق التالية توصي بها لتقدير $\overline{y/x}$ ؟ (ا) استخدم دائمًا $\bar{y}/\bar{X}$ ؟. (ب) استخدم أحيانًا $\bar{y}/\bar{X}$ وأحيانًا $\bar{y}/\bar{X}$ و $\bar{y}/\bar{x}$. (جـ) استخدم دائمًا $\bar{y}/\bar{x}$ ، أعط أسبابًا لجوابك.

(٦-٥) البيان الإحصائي التالي يخص مجتمعًا اصطناعيًا فيه $N=8$ وطبقتين بالحجم نفسه:

طبقة 1		طبقة 2	
x_{1i}	y_{1i}	x_{2i}	y_{2i}
2	0	10	7
5	3	18	15
9	7	21	10
15	10	25	16

قارن تبايني $\hat{Y}_{Rs}$ ، $\hat{Y}_{Rc}$ في عيّنة عشوائية طبقية فيها $n_1=n_2=2$ وذلك بحساب النتائج في جميع العيّنات الممكنة. إلى أي مدى يعود الفرق بين التباينين إلى الانحياز في التقديرات؟

(٦-٦) في التمرين (٦-٥) احسب التباين الناتج عن استخدام طريقة Lahiri في اختيار العيّنة ضمن كل طبقة، واحسب التقدير النسبة المنفصل.

(٦-٧) رُتّبت 45 ولاية من الولايات المتحدة (باستثناء الخمس الأكبر منها) في تسع طبقات تتضمن كل منها خمس ولايات، وللولايات ضمن كل طبقة تقريبًا نسبة عدد سكان 1950 إلى عدد سكان 1940 نفسها. وقد أعطت عيّنة عشوائية طبقية، فيها $n_h=2$ ، النتائج التالية لمجتمع (1960) (y) ، ومجتمع (1950) (x) ، بالملايين.

إذا علمنا أن عدد السكان الإجمالي عام 1950 هو $X=97.94$ فقدر عدد السكان عام 1960 باستخدام التقدير النسبة المركب. أوجد الخطأ المعياري لتقديرك مستخدمًا الطريقة

	طبقة								
	1	2	3	4	5	6	7	8	9
y_{h1}	0.23	0.63	0.97	2.54	4.67	4.32	4.56	1.79	2.18
x_{h1}	0.13	0.50	0.91	2.01	3.93	3.96	4.06	1.91	1.90
y_{h2}	4.95	2.85	0.61	6.07	3.96	1.41	3.57	1.86	1.75
x_{h2}	2.78	2.38	0.53	4.84	3.44	1.33	3.29	2.01	1.32

المختزلة لـ Keyfitz (فقرة ٦-١٣). والعدد الإجمالي الصحيح عام 1960 كان 114.99 هل يتفق تقديرك مع هذا الرقم في حدود أخطاء المعاينة؟

(٦ - ٨) في مثـال التقدير النسبة بمتغيرين الذي أعطاه Olkin سُحبت عيّنة من 50 مدينـة من مجتمـع من 200 مدينـة كبرى. والمتغـيرات هي x_2, x_1, y أعـداد السكـان للمـدينـة الـواحـدة في أعـوام 1950, 1940 و 1930 على الـترتيب. ولـدينـا في المجتمـع، $\bar{Y} = 1699$، $\bar{X}_1 = 1482$، $\bar{X}_2 = 1420$ (بالمئات)، وفي العيّنة، $\bar{y} = 1896, \bar{x}_1 = 1693, \bar{x}_2 = 1643$ والمصفوفة C كما عرفناها في الفقرة (٦-٢٠) هي

	y	x_1	x_2
y	1.213	1.241	1.256
x_1	1.241	1.302	1.335
x_2	1.256	1.335	1.381

قدّر $\bar{Y}$ باستخـدام (ا) متوسط العيّنة، (ب) نسبة عدد السكان عام 1950 إلى عدد السكان عام 1940، (جـ) التقدير النسبة بمتغيرين. احسب الخطأ المعياري المقدّر لكل تقدير.

(٦-٩) برهن أنـه مع طريقـة Midzuno في اختيار العيّنة (فقرة ٦-١٥) يكون احتمال سحب أي عيّنة محددة هو

$$\frac{(n-1)!(N-n)!}{(N-1)!}\,\frac{\hat{\sum}(x_i)}{X}$$

(٦-١٠) في المجتمعـات الصغيرة يكون الحد الرئيس في انحياز $\hat{R}$ في عيّنات عشوائية بسيطة حجمها n من الشكل

$$E(\hat{R}-R)=\frac{b_1(1-f)}{n}=\frac{b_1}{n}-\frac{b_1}{N}$$

حيث لا يعتمد b_1 على N,n وإذا كان $n=mg$ والعيّنة مقسومة عشوائيًا إلى g فئة حجم كل منها m ، لتكن $\hat{R}_i = \sum y / \sum x$ مأخوذة فوق عناصر العيّنة الـ $(n-m)$ الباقية عند حذف الفئة j من العيّنة. بيّن أنه في انحياز التقدير

$$w\hat{R} - (w-1)\hat{R}_i$$

ينعدم كلا الحدين في b_1 إذا كان $w=g[1-(n-m)/N]$.

الفصل السابع

مقدّرات الانحدار

(٧-١) تقدير الانحدار الخطي

تقدير الانحدار الخطي، مثله مثل التقدير النسبة، مصمّم لزيادة الدقة باستخدام المتغير المساعد x_i المرتبط مع y_i وعندما نتأمل العلاقة بين y_i و x_i فقد نجد أنه مع كون العلاقة خطية تقريبًا، إلا أن الخط لا يمر من المبدأ. وهذا يقترح تقديرًا قائمًا على الانحدار الخطي لِـ y_i على x_i بدلاً من نسبة المتغيرين.

لنفرض أننا حصلنا على y_i و x_i لكل وحدة من وحدات العيّنة، وأن متوسط المجتمع x_i وهو $\bar{X}$ معروف، فتقدير الانحدار الخطي لِـ $\bar{Y}$ متوسط المجتمع y_i هو

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x}) \tag{7.1}$$

حيث يرمز الدليل lr إلى الانحدار الخطي (Linear regression) و b تقدر للتغير في y عندما يزداد x بمقدار الواحد. والمنطق وراء هذا التقدير هو أنه إذا كان $\bar{x}$ تحت المعدل فينبغي أن نتوقع كون $\bar{y}$ أيضًا تحت المعدل بمقدار $b(\bar{X} - \bar{x})$ ، وذلك بسبب انحدار y_i على x_i ولتقدير مجموع المجتمع Y نأخذ $\hat{Y}_{lr} = N\bar{y}_{lr}$

وقد استخدم Watson (1937) انحدار مساحة ورقة على وزنها لتقدير متوسط مساحة الأوراق على شجرة. وكانت الطريقة عبارة عن وزن كل الأوراق على الشجرة، ثم تحديد وزن ومساحة كل ورقة من أوراق عيّنة صغيرة. وعندئذ نعدّل متوسط الورقة كما حصلنا عليه من العيّنة، مستخدمين الانحدار على وزن الورقة. وبالطبع فإن الفكرة وراء التطبيق هي أنه يمكن وزن الورقة بسرعة، ولكن تحديد مساحتها يستغرق وقتًا أكبر.

ويوضـح هذا المثـال حالـة عامة تكون فيها تقديرات الانحدار مفيدة جدًّا. فلنفرض أنه يمكننا القيام بتقدير سريع x_i لخاصة ما، وذلك في كل وحدة معاينة، كما نتمكن، بطريقة أكثر تكلفة، تحديد القيمة الصحيحة y_i لهذه الخاصة في وحدات عيّنة عشوائية بسيطة. وعلى سبيل المثال، قد يتمكن خبير جرذان من تقدير عدد الجرذان في كل زقـاق من مسـاحـة معينـة في مدينـة، بعد نظرة سريعة، وعندئذ يحدد بواسطة الأفخاخ، العدد الفعلي للجرذان في كل زقاق من عينة عشوائية بسيطة من الأزقّة. وفي تطبيق آخر وصفه Yates (1960) تمّ بالعين المجردة تقدير الأخشاب في كل قطعة من عدد من قطع الأرض مساحة كل منها $\frac{1}{10}$ فدان، كما تمّ قياس الحجم الفعلي للأخشاب من أجل عيّنة من هذه القطع. وتقدير الانحدار:

$$\bar{y}+b(\bar{X}-\bar{x})$$

يعـدّل متـوسط عيّنـة القياسـات الفعلية من خلال انحـدار القياسـات الفعلية على التقديرات السريعة. ولا حاجة لأن تكون التقديرات السريعة خالية من الانحياز. وإذا كان $x_i-y_i=D$ بحيث إن التقـدير السريع هو، باستثنـاء الانحياز الثـابت D تقـدير صحيح تمامًا، فعندئذ وبأخذ $b=1$ يصبح تقدير الانحدار:

$$\bar{y}+(\bar{X}-\bar{x})=\bar{X}+(\bar{y}-\bar{x})$$ = (متوسط مجتمع التقديرات السريعة) + (تعديل من أجل الانحياز)

وإذا لم نفترض نموذج انحدار خطي فإن لمعرفتنا بخواص تقدير الانحدار الآفاق نفسهـا كمعـرفتنـا بخواص التقدير النسبة. وتقدير الانحدار متّسق بالمعنى البدهي للكلمة، ففي الحقيقة عندما تتألف العيّنة من كامل المجتمع فإن $\bar{x}=\bar{X}$ ، ويصبح تقـدير الانحـدار $\bar{Y}$. وكـما سيُبرهن فإن التقـدير منحاز بصورة عامة، ولكن نسبة الانحياز إلى الخطأ المعياري تصبح صغيرة عندما تكون العيّنة كبيرة. ونمتلك لتباين التقدير علاقة موافقة للعيّنات ذات الحجم الكبير، ولكننا نحتاج إلى معلومات حول توزيع التقـدير في حالـة العيّنات الصغيرة، وحول القيمة المطلوبة لِـ n حتى نتمكن عمليًا من استخدام نتائج العيّنة الكبيرة.

وباختيار مناسب لِـ b نجد أن تقدير الانحدار يتضمن كحالات خاصة كلاًّ من المتوسط لكل وحدة والتقدير النسبة. ومن الواضح أنه إذا أخذنا b مساويًا للصفر فإن

وإذا كان $\bar{y}_{lr}=\bar{y}$ نجد $b=\bar{y}/\bar{x}$

$$\bar{y}_{lr}=\bar{y}+\frac{\bar{y}}{\bar{x}}(\bar{X}-\bar{x})=\frac{\bar{y}}{\bar{x}}\bar{X}=\hat{\bar{Y}}_R \tag{7.2}$$

(٧-٢) تقديرات الانحدار
في حالة قيم محددة سلفًا لِـ b

مع أن b تُقدّر، في معظم التطبيقات، من نتائج عيّنة فمن المنطقي أحيانًا اختيار قيمة لِـ b سلفًا. وعند تكرار المسوح الإحصائية، قد تُظهر لنا حسابات سابقة أن القيم التقديرية لِـ b تبقى إلى حد ما ثابتة من عيّنة إلى أخرى. أو إذا كانت x هي قيمة y في حصر شامل حديث، فقد تقترح علينا معرفتنا العامة بالمجتمع أن b ليس بعيدًا عن الـواحد، وهكذا نختار $b=1$ وبما أن نظرية المعاينة الخاصة بتقديرات الانحدار هي نظرية بسيطة ومفيدة عندما يكون b محددًا سلفًا، فسندرس هذه الحالة أولاً.

نظرية (٧-١)

في المعاينة العشوائية البسيطة التي يكون فيها b_0 ثابتًا محددًا سلفًا يكون تقدير الانحدار الخطي

$$\bar{y}_{lr}=\bar{y}+b_0(\bar{X}-\bar{x})$$

غير منحاز وتباينه:

$$V(\bar{y}_{lr})=\frac{1-f}{n}\frac{\sum_{i=1}^{N}[(y_i-\bar{Y})-b_0(x_i-\bar{X})]^2}{N-1} \tag{7.3}$$

$$=\frac{1-f}{n}(S_y{}^2-2b_0S_{yx}+b_0{}^2S_x{}^2) \tag{7.4}$$

ونلاحظ أننا لا نحتاج إلى أية فروض حول العلاقة بين y و x في مجتمع منته.

برهان

بما أن b_0 ثابت عند تكرار المعاينة فنجد من النظرية (٢-١).

$$E(\bar{y}_{lr}) = E(\bar{y}) + b_0 E(\bar{x} - \bar{X}) = \bar{Y} \qquad (7.5)$$

وفضلاً عن ذلك فإن $\bar{y}_{lr}$ هو متوسط العيّنة للمقادير $y_i - b_0(x_i - \bar{X})$ التي يشكل $\bar{Y}$ متوسط مجتمعها. وبالتالي، واستنادًا إلى النظرية (٢-٢)؛

$$V(\bar{y}_{lr}) = \frac{1-f}{n} \cdot \frac{\sum_{i=1}^{N} [(y_i - \bar{Y}) - b_0(x_i - \bar{X})]^2}{N-1} \qquad (7.6)$$

$$= \frac{1-f}{n}(S_y^2 - 2b_0 S_{yx} + b_0^2 S_x^2) \qquad (7.7)$$

نتيجة

تقدير العيّنة غير المنحاز لـ $V(\bar{y}_{lr})$ هو،

$$v(\bar{y}_{lr}) = \frac{1-f}{n} \frac{\sum_{i=1}^{n} [(y_i - \bar{y}) - b_0(x_i - \bar{x})]^2}{n-1} \qquad (7.8)$$

$$= \frac{1-f}{n}(s_y^2 - 2b_0 s_{yx} + b_0^2 s_x^2) \qquad (7.9)$$

وينتج هذا على الفور، بتطبيق النظرية (٢-٤) على المتغير $(y_i + b_0(x_i - \bar{X})$. والسؤال الطبيعي عند هذه النقطة هو: ما هي أفضل قيمة لـ b_0؟ والجواب معطى في النظرية (٧-٢).

نظرية (٧-٢)

قيمة b_0 التي تجعل $V(\bar{y}_{lr})$ أصغر ما يمكن هي:

$$b_0 = B = \frac{S_{yx}}{S_x^2} = \frac{\sum_{i=1}^{N} (y_i - \bar{Y})(x_i - \bar{X})}{\sum_{i=1}^{N} (x_i - \bar{X})^2} \qquad (7.10)$$

وتدعى معامل الانحدار الخطي لـ y على x في المجتمع المنتهي. ونلاحظ أن B لا يعتمد على خواص أي عيّنة مسحوبة. وبالتالي يمكننا، من الناحية النظرية، تحديدها سلفًا. والتباين الأصغري الناتج هو،

$$V_{min}(\bar{y}_{lr}) = \frac{1-f}{n} S_y^2 (1-\rho^2) \qquad (7.11)$$

حيث ρ قيمة المجتمع لمعامل الارتباط بين y و x .

برهان

لنضع

$$b_0 = B + d = \frac{S_{yx}}{S_x^2} + d \qquad (7.12)$$

في العبارة (7.4) الخاصة بـ $V(\bar{y}_{lr})$ فنجد،

$$V(\bar{y}_{lr}) = \frac{1-f}{n}\left[S_y^2 - 2S_{yx}\left(\frac{S_{yx}}{S_x^2} + d\right) + S_x^2\left(\frac{S_{yx}^2}{S_x^4} + 2d\frac{S_{yx}}{S_x^2} + d^2\right)\right]$$

$$= \frac{1-f}{n}\left[\left(S_y^2 - \frac{S_{yx}^2}{S_x^2}\right) + d^2 S_x^2\right] \qquad (7.13)$$

ومن الواضح أن هذا يصبح أصغر ما يمكن عندما يكون $d=0$ وبما أن $\rho^2 = S_{yx}^2/S_y^2 S_x^2$ ،

$$V_{min}(\bar{y}_{lr}) = \frac{1-f}{n} S_y^2 (1-\rho^2) \qquad (7.14)$$

ويمكن الاستفادة من التحليل ذاته لتبيان مدى إمكانية ابتعاد b_0 عن B دون أن نُصاب بخسارة كبيرة في الدقة .

ومن (7.13) و (7.14) نجد،

$$V(\bar{y}_{lr}) = \frac{1-f}{n}[S_y^2(1-\rho^2) + (b_0 - B)^2 S_x^2] \qquad (7.15)$$

$$= V_{min}(\bar{y}_{lr})\left[1 + \frac{(b_0 - B)^2 S_x^2}{S_y^2(1-\rho^2)}\right] \qquad (7.16)$$

وبما $BS_x = \rho S_y$ فيمكن كتابة هذا على الشكل :

$$V((\bar{y}_{lr}) = V_{min}(\bar{y}_{lr})\left[1 + \left(\frac{b_0}{B} - 1\right)^2 \frac{\rho^2}{(1-\rho^2)}\right] \qquad (7.17)$$

وهكذا إذا أردنا للزيادة النسبية في التباين أن تكون أقل من α فيجب أن يكون،

$$\left|\frac{b_0}{B}-1\right| < \sqrt{\alpha(1-\rho^2)/\rho^2} \tag{7.18}$$

وعلى سبيل المثال، إذا كان $\rho=0.7$ فالزيادة في التباين أقل من 10% ، ($\alpha=0.10$) شريطة أن يكون

$$\left|\frac{b_0}{B}-1\right| < \sqrt{(0.1)(0.51)/(0.49)} = 0.32$$

وتوضح العبارة (7.18) أنه لتأمين زيادة نسبية صغيرة في التباين يجب أن تكون b_0/B قريبة من الواحد، إذا كان ρ مرتفعًا جدًا. إلا أنها يمكن أن تبتعد كثيرًا عن الواحد إذا كانت قيمة ρ معتدلة.

(٧-٣) تقديرات الانحدار
عندما نحسب b من العيّنة

تقترح النظرية (٧-٢) أنه إذا كان لابد من حساب b من العيّنة فمن المرجّح أن يكون تقدير المربعات الدنيا المألوف لِـ B تقديرًا كفؤًا أي

$$b = \frac{\sum_{i=1}^{n} (y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n} (x_i-\bar{x})^2} \tag{7.19}$$

وتلعب نظرية الانحدار الخطي دورًا بارزًا في الطرائقية الإحصائية. والنتائج المعروفة لهذه النظرية ليست مناسبة تمامًا لمسوح العيّنات باعتبارها تتطلب الفرض بأن انحدار y على x في المجتمع المدروس هو انحدار خطي، وأن راسب تباين y حول خط الانحدار يبقى ثابتًا، وأن المجتمع لا نهائي. والخطأ الكبير في تحقّق الشرطين الأولين سيعني إمكانية عدم استخدام تقدير الانحدار الخطي. إلا أنه في المسوح التي يُعتقد فيها أن انحدار y على x خطي تقريبًا، سيكون من المفيد أن نستطيع استخدام $\bar{y}_{lr}$ دون الاضطرار إلى فرض التحقيق المضبوط تمامًا لشرط ثبات راسب التباين.

وبالتالي فإننا نقدم هنا معالجة لا تضع أي فرض حول وجود علاقة محددة بين y_i و x_y. وكما في النظرية المشابهة في حالة التقدير النسبة، نحصل على نتائج صالحة للعيّنات الكبيرة فقط.

وباعتبار b كما عرفناها في (7.19)، يكون مقدّر الانحدار الخطي لِـ $\bar{Y}$ من عيّنات عشوائية بسيطة كما يلي:

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x}) = \bar{y} - b(\bar{x} - \bar{X}) \tag{7.20}$$

وسنبيّن في الفقرة (٧-٧) أن للمقدّر $\bar{y}_{lr}$ كما لِـ $\bar{y}_R$ انحيازًا من مرتبة ــ وعند إيجاد خطأ المعاينة لِـ $\bar{y}_{lr}$ نضع في (7.20) معامل الانحدار في المجتمع B كما عرفناه في (7.10) بدلًا من b المحسوبة من العيّنة. وسنبيّن في النظرية (٧-٣) أن الخطأ المرتكب في هذا التقريب هو من مرتبة $\frac{1}{\sqrt{n}}$ بالنسبة إلى الحدود التي احتفظنا بها. وسندرس أولًا العلاقة بين b و B.

لنُدخل المتغير e_i المعرف بالعلاقة:

$$e_i = y_i - \bar{Y} - B(x_i - \bar{X}) \tag{7.21}$$

فلدينا خاصيتان للمتغيرات e_i هما $\sum^N e_i = 0$ و

$$\sum^N e_i(x_i - \bar{X}) = \sum^N (y_i - \bar{Y})(x_i - \bar{X}) - B\sum^N (x_i - \bar{X})^2 = 0 \tag{7.22}$$

وذلك بالاستناد إلى تعريف B والآن،

$$b = \sum^n y_i(x_i - \bar{x}) / \sum^n (x_i - \bar{x})^2$$

$$= \sum^n [\bar{Y} + B(x_i - \bar{X}) + e_i](x_i - \bar{x}) / \sum (x_i - \bar{x})^2$$

$$= B + \sum^n e_i(x_i - \bar{x}) / \sum^n (x_i - \bar{x})^2 \tag{7.23}$$

والنتيجة التي نحتاجها في النظرية (٧-٣) هي أن $(b-B)$ من مرتبة $\frac{1}{\sqrt{n}}$. ومن النظرية (٢-٣) نجد أن $\sum^n e_i(x_i - \bar{x})/(n-1)$ هو تقدير غير منحاز لِـ $\sum^N e_i(x_i - \bar{X})/(N-1)$ الذي يساوي الصفر وفقًا لِـ6(7.22). وهكذا يتوزع $\sum^n e_i(x_i - \bar{x})/(n-1)$ حول متوسط يساوي الصفر عند

تكرار المعاينة. وبما أننا نعلم أن الخطأ المعياري لتغاير عيّنة هو من مرتبة $\frac{1}{\sqrt{n}}$، فيكون $\frac{\sum^{n} e_i(x_i-\bar{x})}{(n-1)}$ من مرتبة $\frac{1}{\sqrt{n}}$. إلا أن $\sum(x_i-\bar{x})^2/(n-1)=s_x^2$ من مرتبة الواحد، وبالتالي يكون $(b-B)$ وهو كما نرى من (7.23) نسبة هذين المقدارين، من مرتبة $\frac{1}{\sqrt{n}}$.

نظرية (٧-٣)

إذا كان b تقدير المربعات الدنيا لـ B و

$$\bar{y}_{lr}=\bar{y}+b(\bar{X}-\bar{x}) \tag{7.24}$$

فعندئذ نجد في عيّنات عشوائية بسيطة حجمها n كبير أن

$$V(\bar{y}_{lr})\doteq\frac{(1-f)}{n}S_y^2(1-\rho^2) \tag{7.25}$$

حيث $\rho=S_{yx}/S_yS_x$ هو معامل الارتباط في المجتمع بين y و x .

برهان

ينشأ خطأ المعاينة لـ $\bar{y}_{lr}$ من المقدار،

$$\bar{y}_{lr}-\bar{Y}=\bar{y}-\bar{Y}+b(\bar{X}-\bar{x}) \tag{7.26}$$

وكتقريب لنضع بدلاً من $\bar{y}_{lr}$ المقدار،

$$\bar{y}_{lr}=\bar{y}+B(\bar{X}-\bar{x}) \tag{7.27}$$

حيث B معـامـل الانحـدار الخـطي في المجتمـع لـ y على x فالخـطأ المـرتكب في هذا التقريب هو $(B-b)(\bar{X}-\bar{x})$. وهذا المقدار هو من مرتبة في عيّنة عشوائية بسيطة حجمها n باعتبار أن كلاً من $(b-B)$ و $(\bar{x}-\bar{X})$ من مرتبة $\frac{1}{\sqrt{n}}$ ولكن خطأ المعاينة في $\bar{y}_{lr}$ من مرتبة $\frac{1}{\sqrt{n}}$ باعتباره الخطأ في متـوسط العيّنـة للمتغير (y_i-Bx_i) . وبالتالي يكون الحد الرئيس في $E(\bar{y}_{lr}-\bar{Y})^2$ هو $V(\bar{y}_{lr})$. وبالاستناد إلى (7.11) نجد في عيّنات كبيرة،

$$E(\bar{y}_{lr}-\bar{Y})^2\doteq V(\bar{y}_{lr})=\frac{(1-f)}{n}S_y^2(1-\rho^2) \tag{7.28}$$

(٧-٤) تقدير التباين من العيّنة

كتقدير عيّنة لِـ $V(\bar{y}_{lr})$ صالح للتطبيق في العيّنات الكبيرة، يمكن استخدام

$$v(\bar{y}_{lr}) = \frac{1-f}{n(n-2)} \sum_{i=1}^{n} [(y_i - \bar{y}) - b(x_i - \bar{x})]^2 \tag{7.29}$$

$$= \frac{1-f}{n(n-2)} \left\{ \sum (y_i - \bar{y})^2 - \frac{[\sum (y_i - \bar{y})(x_i - \bar{x})]^2}{\sum (x_i - \bar{x})^2} \right\} \tag{7.30}$$

والعلاقة الأخيرة هي العلاقة المختزلة المعتادة في الحسابات. واستنتاجها كما يلي:

لدينا من المعادلة (7.28) في النظرية (٧-٣)، وباعتبار $S_y^2(1-\rho^2) = S_e^2$

$$V(\bar{y}_{lr}) \doteq \frac{(1-f)}{n} S_e^2$$

ومن النظرية (٢-٤)، نجد تقديرًا غير منحاز لِـ S_e^2 هو:

$$s_e^2 = \frac{1}{n-1} \sum_{i=1}^{n} (e_i - \bar{e})^2$$

والآن، نستنتج من المعادلة (7.21) أن،

$$e_i - \bar{e} = (y_i - \bar{y}) - B(x_i - \bar{x}) = [(y_i - \bar{y}) - b(x_i - \bar{x})] + (b - B)(x_i - \bar{x}) \tag{7.31}$$

ويمكن إهمال الحد الثاني على اليمين، وهو من مرتبة $\frac{1}{\sqrt{n}}$ بالنسبة إلى الحد الأول، وهو من مرتبة الواحد. وبالتالي يمكن أن نستخدم في العيّنات الكبيرة

$$\frac{1}{n-1} \sum_{i=1}^{n} [(y_i - \bar{y}) - b(x_i - \bar{x})]^2 \tag{7.32}$$

كتقـدير لِـ S_e^2 وفي (7.29) و (7.30) أقـترح العـامل (n-2) في المقام بدلاً من (n-1) باعتبـاره العامل المستخدم في نظرية الانحدار بشكل عام، ومن المعروف أنه يعطي تقديرًا غير منحاز لِـ S_e^2 إذا كان المجتمع لا نهائيًا والانحدار خطيًّا.

(٧-٥) مقارنة في حالة العيّنات الكبيرة
مع التقدير النسبة ومع المتوسط لكل وحدة

في مثل هذه المقارنات، يجب أن يكون حجم العيّنة كبيراً بكفاية بحيث يصحّ تطبيق العـلاقات التقريبية الخاصة بتباين تقدير الانحدار. والتباينات الثلاثة لتقدير متوسط المجتمع $\bar{Y}$ التي يمكن مقارنتها هي كما يلي:

$$V(\bar{y}_{lr}) = \frac{N-n}{Nn} S_y^2(1-\rho^2) \qquad \text{(انحدار)}$$

$$V(\bar{y}_R) = \frac{N-n}{Nn}(S_y^2 + R^2S_x^2 - 2R\rho S_y S_x) \qquad \text{(نسبة)}$$

$$V(\bar{y}) = \frac{N-n}{Nn} S_y^2 \qquad \text{(المتوسط لكل وحدة)}$$

ومن الواضح أن تباين تقدير الانحدار أصغر من تباين المتوسط لكل وحدة ما لم يكن $\rho = 0$ ، الحالة التي يتساوى فيها التباينان.

وتباين تقدير الانحدار أقل من تباين التقدير النسبة إذا كان:

$$-\rho^2S_y^2 < R^2S_x^2 - 2R\rho S_y S_x \qquad (7.33)$$

وهذا يكافىء المتباينات:

$$(B-R)^2 > 0 \qquad \text{أو} \qquad (\rho S_y - RS_x)^2 > 0 \qquad (7.34)$$

وهكذا يكون تقدير الانحدار أدق من التقدير النسبة ما لم يكن $B=R$. ويحدث هذا عندما تكون العلاقة بين y_i و x_i علاقة خط مستقيم يمر من المبدأ.

مثـال

يمكن مقـارنة تقديرات: الانحدار، النسبة، والمتوسط لكل وحدة، من عيّنة عشـوائية بسيطة، مستخدمين البيان الإحصائي المتوافر من التعداد الكامل لبساتين الدراق الموصوف على الصفحة ٢٥٠ وفي ذلك المثال، y_i هو تقدير إنتاج الدراق في بستان و x_i عدد أشجـار الـدراق في هذا البستان. وسنقارن تقديرين للإنتاج الكلي لـ 256 بستانًا، مأخوذين من عيّنة من 100 بستان. ونشك فيما إذا كانت العيّنة كبيرة مما يكفي

لجعل علاقات التباين قابلة تمامًا للتطبيق، باعتبار أن معامل اختلاف كل من $\overline{y}$ و $\overline{x}$ أعلى، إلى حد ما، من 10 بالمائة، إلا أن المثال سيخدم في توضيح الحسابات. والمعلومات الإحصائية الأساسية هي كما يلي:

$$S_y^2 = 6409 \quad S_{yx} = 4434 \quad S_x^2 = 3898$$
$$R = 1.270 \quad \rho = 0.887 \quad n = 100 \quad N = 256$$

$$V(\hat{Y}_{lr}) = \frac{N(N-n)}{n} S_y^2(1-\rho^2)$$
$$= \frac{(256)(156)}{100}(6409)(1-0.787) = 545,000$$

$$V(\hat{Y}_R) = \frac{N(N-n)}{n}(S_y^2 + R^2 S_x^2 - 2RS_{yx})$$
$$= \frac{(256)(156)}{100}[6409 + (1.613)(3898) - 2(1.270)(4434)]$$
$$= 573,000$$
$$V(\hat{Y}) = \frac{N(N-n)}{n} S_y^2 = 2,559,000$$

وهناك القليل من الاختيار بين التقدير النسبة وتقدير الانحدار، كما قد يكون متوقعًا من طبيعة المتغيرات. والطريقتان كلتاهما تتفوق كثيرًا على المتوسط لكل وحدة.

والنتيجة السابقة حول تفوق تقدير الانحدار لا تصح قطعًا إلا في حالة عيّنات كبيرة، وفي عيّنات صغيرة من مجتمعات فعلية يبدو إنجاز تقدير الانحدار مخيبًا للآمال. وفي ثمانية مجتمعات فعلية من النوع الذي استُخدم فيه التقدير النسبة، وجد Rao (1969) في دراسة «مونت كارلو» أن معدّل النسبة $MSE(\hat{Y}_{lr})/MSE(\hat{Y}_R)$ هو 1.15 في حالة $n=12$ ، و 1.36 عندما $n=8$ و 1.15 عندما $n=6$. وهذه الكفاءة المتدنية لـ $\hat{Y}_{lr}$ لا تعود إلى انحيازات أكبر في تقديرات الانحدار، فالنسبة الموافقة للتباين لها تقريبًا القيمة نفسها.

(٧-٦) دقة علاقات العيّنة

الكبيرة من أجل $V(\bar{y}_{lr})$ و $v(\bar{y}_{lr})$

لا تتوافر أية نتائج تحليلية عامة حول دقة علاقات التقريب (7.25) الخاصة بـ $V(\bar{y}_{lr})$ و (7.29) الخاصة بـ $v(\bar{y}_{lr})$ ، وذلك في حالة عيّنات صغيرة أو معتدلة الحجم. والمقدّرات التقريبية في (7.25) و (7.29) هي ،

$$V(\bar{y}_{lr}) = \frac{(1-f)}{n} S_y^2(1-\rho^2) \tag{7.25}$$

$$v(\bar{y}_{lr}) = \frac{1-f}{n(n-2)} \sum_i^n [(y_i - \bar{y}) - b(x_i - \bar{x})]^2 \tag{7.29}$$

لنفرض أن المتغيرات $i=1,2,\ldots,N, y_i$ هي عيّنة عشوائية من مجتمع لا نهائي خاضع للنموذج ،

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{7.35}$$

حيث تتوزع المتغيرات ε_i مع بقاء x_i مثبتة، مستقلة بعضها عن بعض بمتوسط يساوي الصفر وتباين هو $\sigma_\varepsilon^2 = \sigma_y^2(1-\rho^2)$ ومع هذا النموذج أعطى Cochran (1942) النتيجة التالية وهي أنه إلى حدود من مرتبة $1/n^2$ يكون

$$EV(\bar{y}_{lr}) = \frac{(1-f)}{n} \sigma_y^2(1-\rho^2)\left(1 + \frac{1}{n-3} + \frac{2G_1^2}{n^2}\right) \tag{7.36}$$

حيث $G_1 = k_{3x}/\sigma_x^3$ هو قياس Fisher للالتواء النسبي لتوزيع x وبما أن $S_y^2(1-\rho^2)$ في (2.25) هو، تحت هذا النموذج، تقدير غير منحاز لـ $\sigma_y^2(1-\rho^2)$ فإن (7.36) تقترح، في حالة توزيع متناظر لـ x وتحت هذا النموذج، أن النسبة المئوية للانخفاض في التقدير الناتج عن استخدام $V(\bar{y}_{lr})$ هي $100/(n-2)$.

ومن دراسات مونت كارلو حول ثمانية مجتمعات فعلية صغيرة، (1968, Rao) يبيّن الجدول (٧-١) في حالة $n=6,8,12$ متوسط النسبة المئوية للانخفاض في تقدير تباين $\bar{y}_{lr}$ مستخدمين التقريبين $V(\bar{y}_{lr})$ و $v(\bar{y}_{lr})$.

جدول (٧-١) متوسط النسبة المئوية للانخفاض في تقدير تباين $\bar{y}_{lr}$

المقدّر	n: 6	8	12
$V(\bar{y}_{lr})$ in (7.25) في (7.25)	38	34	28
$v(\bar{y}_{lr})$ in (7.29) في (7.29)	48	42	33

ومن أجل توزيع متناظر لِـ x_i تقترح العلاقة (7.36) انخفاضات في التقدير عند استخـدام $V(\bar{y}_{lr})$ بنسب مئـوية تساوي 25% ، 17% ، و 10% من أجل n تساوي 6 ، 8 ، و 12 على الترتيب. والنسب المئوية ، في الجدول (٧-١) ، الخاصة بِـ $V(\bar{y}_{lr})$ هي أعلى من ذلك بمقادير يصبح معها التعليل من خلال الالتواء (أو عدم التناظر) في قيم x في هذه المجتمعات أمرًا صعبًا والتعليل الأكثر رجحانًا هو أن نقول بأن هذا الارتفاع في النسب يعود إلى نقص في خطيّة النموذج. وتبقى الانخفاضات في التقدير أكبر من أجل تقديرات العيّنة للتباين $v(\bar{y}_{lr})$ وفضلاً عن ذلك، تُظهر المقارنة مع الجدول (٦-٢) في الفصل السابق، والذي ينطبق على التقدير النسبة في المجتمعات الثمانية هذه بالذات، أن النسب المئوية للانخفاضات في التقدير في V و v من أجل $\bar{y}_{lr}$ هي على الأقل ضعف تلك الموافقة لِـ $\bar{y}_R$ في عيّنات من الحجم نفسه.

(٧-٧) الانحياز في تقدير الانحدار الخطي

وللمقدّر $\bar{y}_{lr}$ انحياز من مرتبة $1/n$ في معاينة عشوائية بسيطة فلدينا:

$$E(\bar{y}_{lr}) = \bar{Y} - Eb(\bar{x} - \bar{X}) \quad (7.37)$$

وهكذا تكون إحدى العبارات الممكنة للانحياز هي $-Eb(\bar{x} - \bar{X}) = -\text{cov}\,(b, \bar{x})$. ونجد أن الحد الرئيس في الانحياز هو:

$$\frac{-(1-f)}{n} \frac{Ee_i(x_i - \bar{X})^2}{S_x^2} \quad (7.38)$$

ويمثل هذا الحدّ مساهمة من المركبة التربيعية لانحدار y على x وهكذا إذا بدا الرسم البياني لقيم $\bar{y}_i$ مقابل قيم $\bar{x}_i$ في العيّنة، خطيًا تقريبًا، فينبغي أن تكون المجازفة بوجود

انحياز مهم في $\bar{y}_{lr}$ مجازفة بسيطة .

وبرهان (7.38) يحتاج إلى بعض المعالجة الجبرية . ومن المعادلة (7.23) نعلم أن

$$b = B + \frac{\sum^{n} e_i(x_i - \bar{x})}{\sum^{n} (x_i - \bar{x})^2} \tag{7.39}$$

وإذا وضعنا بدلاً من $\sum^{n} (x_i - \bar{x})^2$ حدها الرئيس $nS_x{}^2$ وكتبنا أيضًا :

$$\sum^{n} e_i(x_i - \bar{x}) = \sum^{n} e_i(x_i - \bar{X}) + n\bar{e}(\bar{X} - \bar{x}) \tag{7.40}$$

فإن الحد الرئيس للانحياز $-Eb(\bar{x} - \bar{X})$ of $\bar{y}_{lr}$ الخاص بِـ $\bar{y}_{lr}$ سيكون متوسط المقادير

$$\frac{-\sum^{n} e_i(x_i - \bar{X})(\bar{x} - \bar{X})}{nS_x{}^2} + \frac{\bar{e}(\bar{x} - \bar{X})^2}{S_x{}^2} \tag{7.41}$$

ليكن $u_i = e_i(x_i - \bar{X})$ فمن (7.38) نجد أن متوسط المجتمع $\bar{U}$ يساوي الصفر. وبالتالي يمكن كتابة القيمة المتوسطة للحد الأول من (7.41) على الشكل ،

$$\frac{-E(\bar{u} - \bar{U})(\bar{x} - \bar{X})}{S_x{}^2} = -\frac{(1-f)}{n} \frac{E(u_i - \bar{U})(x_i - \bar{X})}{S_x{}^2} \tag{7.42}$$

وذلك بالاستناد إلى النظرية (٢-٣) الخاصة بقيمة تغاير عيّنة في معاينة عشوائية بسيطة . وهذا بدوره يساوي (7.38) أي

$$-\frac{(1-f)}{n} \frac{Ee_i(x_i - \bar{X})^2}{S_x{}^2}$$

وفي الحد الثاني من (7.41) ، نجد أن $\bar{e}$ من مرتبة $\frac{1}{\sqrt{n}}$ و $(\bar{x} - \bar{X})^2$ من مرتبة $\frac{1}{n}$ وهكذا يكون هذا الحـد من مرتبـة أدنى من مرتبـة (7.38) وبـالتالي يكون (7.38) الحد الرئيس في انحياز $\bar{y}_{lr}$.

(٧ - ٨) مقدّر الانحدار الخطي تحت نموذج انحدار خطي

لنفرض أن قيم المجتمع المنتهي y_i $(i=1,2,\ldots,N)$ مسحوبة عشوائيا من مجتمع فوقي لانهائي فيه،

$$y = \alpha + \beta x + \varepsilon \tag{7.43}$$

حيث المتغيرات ε مستقلة، ولكل منها متوسط هو الصفر وتباين هو σ_ε^2 وذلك من أجل قيمة مثبتة لِـ x . وبالتعويض المباشر من النموذج نجد أن،

$$b = \frac{\sum^n y_i(x_i - \bar{x})}{\sum^n (x_i - \bar{x})^2} = \beta + \frac{\sum^n \varepsilon_i(x_i - \bar{x})}{\sum^n (x_i - \bar{x})^2} \tag{7.44}$$

$$\bar{y}_{lr} - \bar{Y} = (\bar{\varepsilon}_n - \bar{\varepsilon}_N) + (\bar{X} - \bar{x})\frac{\sum^n \varepsilon_i(x_i - \bar{x})}{\sum^n (x_i - \bar{x})^2} \tag{7.45}$$

حيث $\bar{\varepsilon}_N$ و $\bar{\varepsilon}_n$ متوسطان فوق العيّنة والمجتمع المنتهي، على الترتيب. ونستنتج من (7.45) أن $E(\bar{y}_{lr} - \bar{Y}) = 0$ تحت هذا النموذج، بحيث يكون $\bar{y}_{lr}$ نموذج لا انحياز مهما يكن حجم العيّنة.

وفيما يتعلق بالتبـاين، نستنتـج من (7.45) أنـه من أجـل مجمـوعـة معـطاة من قيم x لدينا:

$$V(\bar{y}_{lr}) = E(\bar{y}_{lr} - \bar{Y})^2 = \sigma_\varepsilon^2\left[\left(\frac{1}{n} - \frac{1}{N}\right) + \frac{(\bar{X} - \bar{x})^2}{\sum^n (x_i - \bar{x})^2}\right] \tag{7.46}$$

وتصحّ هذه النتيجة لأي $n>1$ وأي عيّنة نختارها حصرًا وفقًا لقيم x . وهذا الأسلوب مع تعميمه إلى حالة رواسب تباين غير متساوية كان من إنتاج Royall (1970) . وتحت هذا النموذج فإن خطة المعاينة الهادفة، التي تنجح في جعل $\bar{x}=\bar{X}$ ستجعل $V(\bar{y}_{lr})$ أصغر مايمكن من أجل قيمة معطاة لِـ n . وأيضًا، من أجل أي عيّنة نختارها حصرًا وفقًا للقيم x_i يكون مقدر المربعات الدنيا المعتاد،

$$s_e^2 = \sum^n [(y_i - \bar{y}) - b(x_i - \bar{x})]^2/(n-2) \tag{7.47}$$

مقدّر نموذج ـ لا انحياز لـ σ_ε^2 وذلك عندما $n>2$.

وهكذا، ففي المسائل التي ينطبق فيها هذا النموذج يمكن إرساء نتائج بسيطة ومضبوطة حول متوسط وتباين $\bar{y}_{lr}$ ، ويصحّ تطبيقها لأي حجم للعيّنة يتجاوز 2 ، ولا تتطلب إلا اختيار العيّنة وفقًا للقيم x_i أما عنصر العشوائية فيمدّنا به مجانًا توزيع المقادير ε المفترضة في النموذج.

(٧ـ٩) تقديرات الانحدار في معاينة طبقية

وكما في حالة التقدير النسبة هناك نوعان من تقديرات الانحدار يمكن القيام بهما في معاينة عشوائية طبقية. ففي التقدير الأول $\bar{y}_{lrs}$ (s من أجل separate)، نحسب تقدير الانحدار لمتوسط كل طبقة، أي نأخذ،

$$\bar{y}_{lrh} = \bar{y}_h + b_h(\bar{X}_h - \bar{x}_h) \tag{7.48}$$

وعندئذ نأخذ،

$$\bar{y}_{lrs} = \sum_h W_h \bar{y}_{lrh} \tag{7.49}$$

حيث $W_h=N_h/N$. ويكون هذا التقدير مناسبًا عندما يبدو لنا أن معامل الانحدار الحقيقي يتغير من طبقة إلى طبقة.

أما تقدير الانحدار الثاني $\bar{y}_{lrc}$ (c من أجل combined) فيكون مناسبًا عندما يُفترض أن B_h يبقى نفسه في جميع الطبقات. ولحساب $\bar{y}_{lrc}$ نجد أولاً،

$$\bar{y}_{st} = \sum_h W_h \bar{y}_h \qquad \bar{x}_{st} = \sum_h W_h \bar{x}_h$$

وعندئذ يكون،

$$\bar{y}_{lrc} = \bar{y}_{st} + b(\bar{X} - \bar{x}_{st}) \tag{7.50}$$

وسندرس التقديرين أولاً في الحالة التي نختار فيها b_h و b سلفًا، باعتبار أن خواصها تكون بسيطة بصورة غير عادية في هذه الحالة. ومن الفقرة (٧ـ٢) نجد أن $\bar{y}_{lrh}$ تقدير غير منحاز لـ $\bar{Y}$ أي أن $\bar{y}_{lrs}$ تقدير غير منحاز لـ $\bar{Y}$. وبما أن المعاينة مستقلة في الطبقات المختلفة فنستنتج من النظرية (٧ـ١) أن،

$$V((\bar{y}_{lrs}) = \sum_h \frac{W_h{}^2(1-f_h)}{n_h}(S_{yh}{}^2 - 2b_h S_{yxh} + b_h{}^2 S_{xh}{}^2) \tag{7.51}$$

وتبينّ النظرية (٧-٢) أن $V(\bar{y}_{lrs})$ يكون أصغر ما يمكن عندما يكون b_h مساويًا لمعامل الانحدار الحقيقي B_h في الطبقة h ويمكن كتابة القيمة الصغرى للتباين على الشكل،

$$V_{min}(\bar{y}_{lrs}) = \sum_h \frac{W_h{}^2(1-f_h)}{n_h}\left(S_{yh}{}^2 - \frac{S_{yxh}^2}{S_{xh}{}^2}\right) \tag{7.52}$$

واذا التفتنـــا الآن إلى التقــدير المــركب، مع b محددة سلفًــا، يتبـينّ لنـا، من (7.50) أن $\bar{y}_{lrc}$ هو أيضًا تقدير غير منحاز لـ $\bar{Y}$ في هذه الحالة. وبما أن $\bar{y}_{lrc}$ هو التقدير المعتاد من عينة طبقية للمتغير $y_{hi} + b(\bar{X} - x_{hi})$، يمكن تطبيق النظرية (٥-٣) على هذا المتغير لنجد النتيجة التالية:

$$V(\bar{y}_{lrc}) = \sum_h \frac{W_h{}^2(1-f_h)}{n_h}(S_{yh}{}^2 - 2bS_{yxh} + b^2S_{xh}{}^2) \tag{7.53}$$

وقيمة b التي تجعل التباين أصغر ما يمكن هي،

$$B_c = \sum_h \frac{W_h{}^2(1-f_h)S_{yxh}}{n_h} \Big/ \sum_h \frac{W_h{}^2(1-f_h)S_{xh}{}^2}{n_h} \tag{7.54}$$

والمقدار B_c هو متوسط مرجّح لمعاملات الانحدار الطبقية $B_h = S_{yxh}/S_{xh}{}^2$ وإذا كتبنا

$$a_h = \frac{W_h{}^2(1-f_h)}{n_h} S_{xh}{}^2$$

فعندئذ يكون $B_c = \Sigma\, a_h B_h / \Sigma\, a_h$.

ومن (7.52) و (7.53) بعد وضع B_c بدلاً من b نجد،

$$\begin{aligned} V_{min}(\bar{y}_{lrc}) - V_{min}(\bar{y}_{lrs}) &= \sum a_h B_h{}^2 - (\sum a_h)\, V_c{}^2 \\ &= \sum a_h (B_h - B_c)^2 \end{aligned} \tag{7.55}$$

وتبينّ هذه النتيجة أنه في حالة الاختيارات المثلى للتقدير المنفصل سيكون تباينه أصغر من تباين التقدير المركب ما لم يكن B_h نفسه في جميع الطبقات. وسوف تستدعي هذه الاختيارات المثلى، بالطبع، معرفة مسبقة بقيم الـ S_{yxh} و $S_{xh}{}^2$.

(٧-١٠) معاملات انحدار مقدّرة من العيّنة

يكـون التحليل السـابق مفيدًا للدلالـة على نوع تقـديرات العيّنـة b_h ونـوع المقادير b التي يمكن أن تكون فعّالة عند استخدامها في تقديرات الانحدار. وفي حالة تقدير منفصل، يقترح هذا التحليل أخذ تقدير المربعات الدنيا لِـ B_h ضمن الطبقة وهو،

$$b_h = \frac{\sum_i (y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h)}{\sum_i (x_{hi} - \bar{x}_h)^2} \tag{7.56}$$

وبتطبيق النظرية (٧-٣) على كل طبقة نجد،

$$V(\bar{y}_{lrs}) = \sum_h \frac{W_h^2(1-f_h)}{n_h} S_{yh}^2(1-\rho_h^2) \tag{7.57}$$

شريطة أن يكـون حجم العيّنة كبيرًا في جميع الطبقات. وللحصول على تقدير عيّنة للتباين نعوض

$$s_{y \cdot xh}^2 = \frac{1}{n_h - 2}\left[\sum_i (y_{hi} - \bar{y}_h)^2 - b_h^2 \sum_i (x_{hi} - \bar{x}_h)^2\right] \tag{7.58}$$

بدلاً من $S_{yh}^2(1-\rho_h^2)$ في (7.57) .

ويعاني التقدير $\bar{y}_{lrs}$ من الصعوبات نفسها التي يعاني منها التقدير النسبة الموافق، من حيث إنه قد تصبح نسبة انحياز $\bar{y}_{lrs}$ إلى خطئه المعياري غير قليلة. ونستنتج من الفقرة (٧-٧) أن انحياز تقديرات الانحدار $\bar{y}_{lrs}$ في كل طبقة بمفردها يمكن أن يكون من مرتبة $\frac{1}{n_h}$ وأن الانحيازات قد تكون من الإشارة نفسها في جميع الطبقات، بحيث يصبح الانحياز الكلي في $\bar{y}_{lrs}$ أيضًا من مرتبة $\frac{1}{n_h}$ وبما أن الحد الرئيس في الانحياز يأتي من الانحدار التربيعي لِـ y_{hi} على x_{hi} كما بينّا في الفقرة (٧-٧)، فإن هذا الخطر يكون أكثر حدّة عندما تشكل العلاقة بين المتغيرات تقريبًا لعلاقة من النوع التربيعي بدلاً من النوع الخطي.

وفي حالة تقدير مركب، رأينا أن التباين يصبح أصغريًا عندما يكون $b=B_c$ كما عرفناه في (7.54) ، وهذا يقترح أخذ،

$$b_c = \sum_h \frac{W_h^2(1-f_h)}{n_h(n_h-1)} \sum_i (y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h) \Big/ \sum_h \frac{W_h^2(1-f_h)}{n_h(n_h-1)} \sum_i (x_{hi} - \bar{x}_h)^2$$

كتقدير عيّنة لِـ B_c وإذا كان التقسيم إلى طبقات متناسبًا وأمكن أن نضع n_h في عبارة b_c بدلاً من (n_h-1) فإن b_c يأخذ الصيغة المألوفة لتقدير المربعات الدنيا المركب:

$$b_c' = \sum_h \sum_i (y_{hi} - \bar{y}_h)(x_{hi} - \bar{x}_h) \Big/ \sum_h \sum_i (x_{hi} - \bar{x}_h)^2$$

وفي ظروف معينة يمكن أن نفضل تقديرات أخرى على b_c أو b'_c وعلى سبيل المثال، إذا كانت معاملات الانحدار الصحيحة B_h نفسها في جميع الطبقات، إلا أن رواسب التباينات حول خط الانحدار تختلف بحدّة من طبقة إلى أخرى، فقد يكون متوسط مرجّح آخر للمقادير b_h ، نرجّح فيه وفق مقلوب تقدير التباين، أكثر دقة. وعلى أي حال، فمن المرجّح أن يكون الكسب في الدقة فيما يتعلق بِـ $\bar{y}_{lrc}$ صغيرًا.

وبما أن،

$$\bar{y}_{lrc} - \bar{Y} = \bar{y}_{st} - \bar{Y} + b_c(\bar{X} - \bar{x}_{st})$$

$$= [\bar{y}_{st} - \bar{Y} + B_c(\bar{X} - \bar{x}_{st})] + (b_c - B_c)(\bar{X} - \bar{x}_{st}) \qquad (7.59)$$

فنستنتج، في حال إمكانية إغفال أخطاء المعاينة في b_c ، أن،

$$V(\bar{y}_{lrc}) = \sum_h \frac{W_h^2(1-f_h)}{n_h}(S_{yh}^2 - 2B_c S_{yxh} + B_c^2 S_{xh}^2) \qquad (7.60)$$

وكتقدير لِـ $V(\bar{y}_{lrc})$، يمكن أخذ،

$$v(\bar{y}_{lrc}) = \sum_h \frac{W_h^2(1-f_h)}{n_h(n_h-1)} \sum_i [(y_{hi} - \bar{y}_h) - b_c(x_{hc} - \bar{x}_h)]^2 \qquad (7.61)$$

(٧-١١) مقارنة نوعي تقديرات الانحدار

لا يمكن إعطاء قواعد سريعة وحاسمة لتقدير ما إذا كان التقدير المنفصل أو المركب هو الأفضل في أي حالة محددة. وعيوب التقدير المنفصل، هي أنه أكثر عرضة للانحياز عندما تكون العيّنات صغيرة ضمن كل طبقة بمفردها، وأن الإسهام الأكبر

في تباينه يأتي من أخطاء المعاينة في معاملات الانحدار. وعيب التقدير المركب هو أن تباينه يتضخم إذا اختلفت معاملات الانحدار في المجتمع من طبقة إلى أخرى.

وإذا كنا على ثقة بأن الانحدارات خطية وبدت المقادير B_h وكأنها تبقى نفسها تقريبًا في جميع الطبقات، فالأفضلية ينبغي أن تكون للتقدير المركب. أما إذا بدا أن الانحـدارات خطية (بحيث يبـدو خطر الانحياز صغـيرًا) إلا أن المقادير B_h تختلف اختلافًا ملحوظًا من طبقة إلى أخرى، فيُوصَى بالتقدير المنفصل. وفي حال وجود بعض الانحنـاء في الانحـدار بينما نستخدم تقدير انحدار خطي، فمن المحتمل أن يكون التقدير المركب أكثر أمانًا، هذا ما لم تكن العيّنات كبيرة في جميع الطبقات.

وكـان ميكي Mickey (1959) ، وويليام William (1963) قد ابتكـرا مقدرات الانحدار غير المنحازة إلا أنها لم تُجرّب على نطاق واسع بعد. وقد وجد Rao (1969) أن مقدّر Mickey متخلف عادة عن مقدّر الانحدار والمقدّر النسبة المعتادين في مجتمعات فعلية. ويمكن أيضًا اتباع أسلوب مدية الجيب. ففي حالة $n=mg$ ليكن $\bar{y}'_{(lr)i}$ تقدير الانحدار المعتاد، محسوبًا من العيّنة بعد أن نحذف منها الفئة j $(j=1,2,\ldots,g)$. فعندئذ يكون تقدير مدية الجيب من الشكل:

$$\bar{y}_{(lr)Q} = g\bar{y}_{lr} - (g-1)\left(\sum^{g} \bar{y}'_{(lr)i}\right) \Big/ g \tag{7.62}$$

تماريـن

(٧-١) يقوم مزارع متمرّس بتقدير وزن الدراق x_i بالعين المجردة في كل شجرة من بستان يحوي $N=200$ شجرة. ويجد أن الوزن الكلي التقديري هو $X=11600$ ليبرة. وقد التُقط الدراق ووُزن في عينة عشوائية بسيطة من عشر شجرات، وكانت النتائج كما يلي:

		رقم الشجرة										
		1	2	3	4	5	6	7	8	9	10	Total
الوزن الفعلي	y_i	61	42	50	58	67	45	39	57	71	53	543
الوزن المقدّر	x_i	59	47	52	60	67	48	44	58	76	58	569

وكتقدير للوزن الفعلي الإجمالي Y ، نأخذ،

$$\hat{Y} = N[\bar{X} + (\bar{y} - \bar{x})]$$

احسب التقدير وأوجد خطأه، المعياري .

(٧-٢) هل يبدو لك أن تقدير الانحدار الخطي ، مع b محسوبة من العيّنة وفقًا لطريقة المربعات الدنيا سيعطي تقديرًا أكثر دقة في (7.1) ؟

(٧-٣) من بيان العيّنـة في الجدول (٦-١) من الفصل السابق احسب تقدير الانحدار لمجموع عدد السكان عام 1930 في الـ 196 مدينة كبرى . أوجد الخطأ المعياري التقريبي لهذا التقدير وقارن دقته مع دقة التقدير النسبة .

(٧-٤) في التمرين (٧-٣) أوجد تقدير العدد الإجمالي للسكان وخطأه المعياري إذا أخذنا b مساويًا للواحد .

(٧-٥) في المجتمع التالي حيث $N=5$ تحقق من (ا) أن انحدار y على x خطي و(ب) أن تقـدير الانحـدار الخـطي غير منحـاز في عيّنات عشوائية بسيطة حجمها $n=3$ والأزواج (y,x) هي ، (3.0), (5,0) (8,2), (8,3), (12,3) .

(٧-٦) العـلاقـة بين قياس تقريبي x ، قمنا به في كل وحدة، وبين القياس الصحيح y في الوحدة، معطاة بالمعادلة

$$x = y + e + d$$

حيث d انحياز ثابت و e خطأ القياس الذي لا يرتبط بـ y وله متوسط يساوي الصفر وتباين في مجتمع نفرضه لا نهائي . في حالة عيّنات عشوائية بسيطة حجم كل منهـا n قارن تبايني : (ا) تقدير الفرق $[\bar{y}+(X-\bar{x})]$ للمتوسط $\bar{Y}$ و(ب) تقدير الانحدار الخطي مستخدمًا قيمة b التي تجعل التباين أصغريًا . (قد تعتمد التباينات على S_y^2.) ،

(٧ـ٧) بأخـذ كل الحالات الممكنة، قارن متوسطي مربعات الخطأ لتقديري الانحـدار المنفصـل والمركب للمجموع الكلي Y للمجتمـع التالـي y ، وذلك عندما نسحب عيّنات عشوائية بسيطة حجمها 2 من كل طبقة. ومن أجل كل تقدير كم تكون مساهمة انحيازه في متوسط مربعات الخطأ؟

طبقة 1		طبقة 2	
x_{1i}	y_{1i}	x_{2i}	y_{2i}
4	0	5	7
6	3	6	12
7	5	8	13

استخدم تقديرات المربعات الدنيا المعتادة للمقادير b_c ، b_h و $B's$ المذكورة في الفقرة (٧ـ١٠).

(٧ ـ ٨) في المجتمع المذكور في التمرين (٧ـ٧)، بيّن أنه إذا أمكن استخدام B المثلى المحددة سلفًا في كل حالة، فعندئـذ $V(\hat{Y}_{lrs}) = 4.39$, $V(\hat{Y}_{lrc}) = 4.43$ وكلا التقديرين، بالطبع، غير منحاز.

(٧ـ٩) بالطريقة نفسها قارن متوسطي مربعات الخطأ للمقدِّرين النسبة المنفصل والمركب في المجتمع المذكور في التمرين (٧ـ٧). وبما أن النسبة Y/X تساوي 8/17=0.47 في الطبقة الأولى و 23/19=1.68 في الطبقة الثانية، فستقترح نظرية العيّنات الكبيرة إمكانية تفوق $\hat{Y}_{Rs}$ على $\hat{Y}_{Rc}$. إلا أنك ستجد في هذه العيّنات الصغيرة جدًّا أن متوسط مربعات الخطأ أصغر من أجل $\hat{Y}_{Rc}$. وتفوقه لا يعود إلى انحياز أصغر فليس لـ $\hat{Y}_{Rs}$ ولا لـ $\hat{Y}_{Rc}$ انحياز يُذكر. وكخلاف آخر مع نظرية العيّنات الكبيرة ستجد أن متـوسـطي مربعـات الخـطأ لكـل من $\hat{Y}_{Rs}$ و $\hat{Y}_{Rc}$ أصغر من متوسطي مربعات الخطأ لتقديري الانحدار الموافقين.

الفصل الثامن

المعاينة النمطيّة

(٨ - ١) وصـــف

هذه الـطريقة في المعاينة مختلفة، من النظرة الأولى، اختلافًا تامًّا عن المعاينة العشـوائية البسيطة. لنفـرض أننـا رقّمنـا الوحدات الـ N في المجتمع، بترتيب ما، من 1 إلى N . فلاختيار عيّنة من n من الوحدات، نأخذ وحدة من الوحدات الـ k الأولى بصـورة عشـوائية، ثم نختار بعدها بصورة نمطية مرتّبة الوحدة الـ k من كل k من الـوحـدات التالية. فمثلاً إذا كان $k=15$ وكانت الوحدة الأولى المسحوبة هي ذات الـرقم 13 فإن الوحدات التالية في العيّنة تكون ذوات الأرقام 28 ، 43 و 58 وهكذا. واختيار الوحدة الأولى يحدد كامل العيّنة. وسيدعى هذا النوع من العيّنة بالعيّنة النمطية كل k وحدة.

والميزات الظاهرة لهذه الطريقة فوق المعاينة العشوائية البسيطة هي كما يلي:

١ - سحب العيّنة أسهل، وغالبًا ما يكون من الأسهل تنفيذ عملية السحب بدون أخطاء. ولهذه الميزة أهميتها عندما يتم السحب في الميدان. وقد يوجد توفير كبير في الـوقت حتى عنـدمـا يتم السحب في مكتب. وعـلى سبيل المثـال، إذا كانت الوحدات موصوفة على بطاقات لها جميعها الحجم نفسه وواقعة في جرّار للأضابير، فيمكن، وعلى طول البطاقات المنضّدة، سحب بطاقة بعد كل بوصة وذلك وفقًا للقياس بمسـطرة مدرجـة. وهـذه العملية سريعة بينما يمكن أن تكون المعاينة العشـوائية البسيطة بطيئة. وبالطبع فإن هذه الطريقة تبتعد قليلاً عن القاعدة الدقيقة «كل k وحدة».

٢ ـ بالبداهة، يبدو من المرجّح أن تكون المعاينة النمطية أكثر دقة من المعاينة العشوائية البسيطة. وفي الحقيقة فإنها تقسم المجتمع طبقيًا إلى n من الطبقات التي تتألف من الوحدات الـ k الأولى، الوحدات الـ k التي تليها، وهكذا. ولذلك يمكننا أن نتوقع كون العينة النمطية في حوالي دقة العيّنة العشوائية الطبقية الموافقة نفسها، مع وحدة واحدة من كل طبقة. والفرق هو أنه في حالة العيّنة النمطية تقع كل الوحدات في الموضع النسبي نفسه في الطبقة. بينما يتحدد الموضع ضمن كل طبقة بصورة منفصلة وعشوائية في المعاينة العشوائية الطبقية (انظر الشكل ٨ـ١). وتنتشر العيّنة النمطية بعدالة أكثر فوق المجتمع، وهذه الحقيقة تجعل المعاينة النمطية أحيانًا أكثر دقة بكثير من المعاينة العشوائية الطبقية.

عينة نمطية = × عينة عشوائية طبقية = o

k $2k$ $3k$ $4k$ $5k$ $6k$

رقم الوحدة

شكل (٨ ـ ١) المعاينة النمطية والمعاينة العشوائية الطبقية

جدول (٨ ـ ١) العيّنات النمطية الممكنة من أجل $k=5, N=23$

رقم العيّنة النمطية

I	II	III	IV	V
1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23		

وأحد الأشكال الأخرى للمعاينة النمطية هو أن نختار كل وحدة عند أو قرب مركز الطبقة. أي بدلًا من أن تبدأ المتوالية بعدد نختاره عشوائيًا بين 1 و k نأخذ نقطة البداية $(k+1)/2$ إذا كان k فرديًا، وإما $k/2$ أو $(k+2)/2$ إذا كان k زوجيًا، Madow (1953)، وينقل هذا الإجراء فكرة المعاينة النمطية إلى خاتمتها المنطقية. وإذا أمكن اعتبار y_i دالة مستمرة في متغير مستمر i، فهناك ما يدعو للتوقع بأن هذه العينة

المتمـوضعـة مركزيًا ستكون أكثر دقة من العيّنة التي تتموضع عشوائيًا. والقليل من التحـرّيات التي تنـاولت مجتمعات فعلية تدعم هذا الرأي، بالرغم من أن العيّنات المتموضعة مركزيًا تنزع الى أن تسلك سلوكًا شاذًا، وسنقصر انتباهنا هنا على العيّنات التي تتضمن عنصرًا عشوائيًا ما.

وبـما أن N ليست، بصورة عامة، من مضاعفات k فيمكن أن تختلف العيّنات النمطية من المجتمع المنتهي نفسه بوحدة واحدة من حيث حجمها. وهكذا فإنه في حالة $k=5, N=23$ تكون أعداد الوحدات في العيّنات النمطية الخمس كما هو مبيّن في الجـدول (٨ ـ ١)، والعيّنـات الثـلاث الأولى حجمها $n=5$، بينما $n=4$ في العيّنتين الأخـيرتين. وهذه الحقيقة تسبب خللاً في نظرية المعاينة النمطية، ومن المحتمل أن يكون هذا الخلل مهملاً إذا تجاوز n الخمسين، وسنتجاهله عند تقديم النظرية توخيًا للبساطة. ومن المرجّح ألا يكون الخلل كبيرًا حتى عندما يكون n صغيرًا.

وقد اقترح Lahiri (1952) (انظر Murthy ,1967 ، ص*139*) طريقة أخرى تزوّدنا بحجم عيّنـة ثابت، ومتـوسط عيّنـة غير منحاز، في الوقت نفسه. لنعتبر الوحدات الـ N وكأنها مرتبة على محيط دائرة وليكن k الآن العدد الصحيح الأقرب لِـ N/n . لنختر عددًا عشـوائيًا بين 1 و N ، ثم لنأخذ الوحدة الـ k من كل k من الوحدات التالية، ماضين هكذا على طول محيط الدائرة حتى يتم اختيار الوحدات الـ n التي نريدها. ولنفرض أننا نريد $n=5$ حيث $N=23$ فعندئذ $k=5$ ، وإذا كان العدد العشوائي 19 نأخذ الـوحـدات 19 ,1 ,6 ,11 ,16 . ومن السهل التحقق أن لكل وحدة في هذه الطريقة الاحتمال نفسه في أن تكون هي الوحدة المختارة. وإذا كان $n=4$ هو الحجم المرغوب حيث $N=23$ فإننا نأخذ $k=6$.

(٨ ـ ٢) الصلة بالمعاينة العنقودية

هناك طريقة أخرى للنظر إلى المعاينة النمطية فمع $N=nk$ تبيّن أعمدة الجدول (٨ ـ ٢) العيّنـات النمـطية الممكنة. ويتضح من هذا الجدول أن المجتمع قد قُسم إلى k من وحدات المعاينة الكبيرة، وكل منها تحوي n من الوحدات الأصلية. وعملية

اختيار عيّنة نمطية متموضعة عشوائيًا هو مجرد اختيار واحدة من وحدات المعاينة الكبيرة هذه بصورة عشوائية. وهكذا فإن المعاينة النمطية تؤدي أساسًا إلى اختيار وحدة معاينة مركبة لتشكل بمفردها مجمل العيّنة. والعيّنة النمطية هي عيّنة عشوائية بسيطة تتضمن وحدة عنقودية واحدة من مجتمع يتضمن k من الوحدات العنقودية.

جدول (٨ ـ ٢) إنشاء k من العيّنات النمطية

	رقم العيّنة					
	1	2	. . .	i	. . .	k
	y_1	y_2		y_i		y_k
	y_{k+1}	y_{k+2}		y_{k+i}		y_{2k}
	. . .	. . .		. . .		. . .
	$y_{(n-1)k+1}$	$y_{(n-1)k+2}$		$y_{(n-1)k+i}$		y_{nk}
المتوسطات	$\bar{y}_1$	$\bar{y}_2$		$\bar{y}_i$		$\bar{y}_k$

(٨ ـ ٣) تباين تقدير متوسط

قد طُوّرت عدة صيغ تتعلق بتباين $\bar{y}_{sy}$ ، متوسط عيّنة نمطية. والصيغ الثلاث المعطاة أدناه تنطبق على أي نوع من المعاينة العنقودية يحوي فيها كل من العناقيد n عنصرًا، وتتألف العيّنة من عنقود واحد. وفي هذه الحالات نفترض أن $N=nk$.

وإذا كان $N=nk$ ، فمن السهل التحقق من أن $\bar{y}_{sy}$ هو تقدير غير منحاز لـ $\bar{Y}$ من أجل عيّنة نمطية عشوائية التموضع.

وفي التحليل التالي يدلّ الرمز y_{ij} على العنصر j من العيّنة النمطية i ، حيث $j=1,2,\ldots,n$ ، $i=1,2,\ldots,k$. ونرمز لمتوسط العيّنة i بـ $\bar{y}_i$.

نظرية (٨ ـ ١)

تباين متوسط العيّنة النمطية هو:

$$V(\bar{y}_{sy})=\frac{N-1}{N}S^2-\frac{k(n-1)}{N}S_{wsy}^2 \qquad (8.1)$$

حيث

$$S_{wsy}^2=\frac{1}{k(n-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$$

هو التباين فيما بين الوحدات التي تقع ضمن العيّنة النمطية نفسها. ونشكل مقام هذا التبـاين، $k(n\text{-}1)$ ، وفقًـا للقواعد المعتادة في تحليل التباين: تُسهم كل من العيّنات الـ k بِـ $(n\text{-}1)$ درجة من الحرية في مجموع المربعات الموجود في البسط.

برهان

بالاستناد إلى المطابقة المعتادة في تحليل التباين لدينا:

$$(N-1)S^2 = \sum_i \sum_j (y_{ij} - \bar{Y})^2$$
$$= n \sum_i (\bar{y}_{i.} - \bar{Y})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$

ولكن تباين $\bar{y}_{sy}$ هو بالتعريف،

$$V(\bar{y}_{sy}) = \frac{1}{k} \sum_{i=1}^{k} (\bar{y}_{i.} - \bar{Y})^2$$

ومنه،

$$(N-1)S^2 = nkV(\bar{y}_{sy}) + k(n-1)S^2_{wsy} \qquad (8.2)$$

وهو المطلوب.

نتيجة

يكون متوسط عيّنة نمطية أكثر دقة من متوسط عيّنة عشوائية بسيطة إذا وفقط إذا كان،

$$S^2_{wsy} > S^2$$

برهان

إذا كان $\bar{y}$ متوسط عيّنة عشوائية بسيطة حجمها n ، فإن،

$$V(\bar{y}) = \frac{N-n}{N} \frac{S^2}{n}$$

ومن المعادلة (8.1) نجد أن $V(\bar{y}_{sy}) < V(\bar{y})$ إذا وفقط إذا كان،

$$\frac{N-1}{N} S^2 - \frac{k(n-1)}{N} S^2_{wsy} < \frac{N-n}{N} \frac{S^2}{n} \qquad (8.3)$$

أي إذا كان،

$$k(n-1)S_{wsy}^2 > \left(N-1-\frac{N-n}{n}\right)S^2 = k(n-1)S^2 \tag{8.4}$$

وهذه النتيجة المهمة التي تنطبق على المعاينة العنقودية، بصورة عامة، تفيد بأن المعاينة النمطية أكثر دقة من المعاينة العشوائية البسيطة، إذا كان التباين ضمن العيّنات النمـطية أكـبر من تبـاين المجتمع ككل. وتكون المعاينة النمطية دقيقة عندما تكون الـوحـدات ضمن العيّنـة نفسها غير متجانسة، وغير دقيقة عندما تكون متجانسة. والنتيجة واضحة بالبداهة. وإذا كان هناك القليل من التغير ضمن عينة نمطية بالنسبة للمتغير المدروس في المجتمع، فإن الوحدات المتتالية في العينة تكرر، إلى حدّ كبير أو صغير، المعلومات نفسها.

وتعطي النظرية (٨ ـ ٢) شكلاً آخر للتباين:

نظرية (٨ ـ ٢)

$$V(\bar{y}_{sy}) = \frac{S^2}{n}\left(\frac{N-1}{N}\right)[1+(n-1)\rho_w] \tag{8.5}$$

حيث ρ_w معـامل الارتباط بين أزواج من الوحدات الموجودة في العيّنة النمطية نفسها. ونعرّفه على الشكل،

$$\rho_w = \frac{E(y_{ij}-\bar{Y})(y_{iu}-\bar{Y})}{E(y_{ij}-\bar{Y})^2} \tag{8.6}$$

حيث البسط هو المتوسط فوق جميع الـ $kn(n-1)/2$ من الأزواج المتميزة.

والمقام هو المتوسط فوق جميع القيم الـ N لِـ y_{ij}. وبما أن المقام هو $(N-1)S^2$ فنجد،

$$\rho_w = \frac{2}{(n-1)(N-1)S^2}\sum_{i=1}^{k}\sum_{j<u}(y_{ij}-\bar{Y})(y_{iu}-\bar{Y}) \tag{8.7}$$

برهان

$$n^2kV(\bar{y}_{sy}) = n^2\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{Y})^2$$

$$= \sum_{i=1}^{k}[(y_{i1}-\bar{Y})+(y_{i2}-\bar{Y})+\cdots+(y_{in}-\bar{Y})]^2$$

ومجموع الحدود المربعة هو مجموع مربعات الانحرافات عن $\bar{Y}$ ، أي أنه يساوي $(N\text{-}1)S^2$. وهذا يعطي ،

$$n^2kV(\bar{y}_{sy}) = (N-1)S^2 + 2\sum_i \sum_{j<u} (y_{ij} - \bar{Y}(y_{iu} - \bar{Y}) \quad (8.8)$$

$$= (N-1)S^2 + (n-1)(N-1)S^2\rho_w \quad (8.9)$$

ومنه ،

$$V(\bar{y}_{sy}) = \frac{S^2}{n}\left(\frac{N-1}{N}\right)[1 + (n-1)\rho_w] \quad (8.10)$$

وهذا يبيّن أن الارتباط الإيجابي بين وحدات العيّنة نفسها يضخّم تباين متوسط العيّنة. وقد يكون حتى لارتباط إيجابي صغير تأثير كبير بسبب العامل $(n\text{-}1)$.

وتعبر النظريتان السابقتان عن $V(\bar{y}_{sy})$ بدلالة S^2 ، وبالتالي تربطانه بالتباين الموافق لعيّنة عشوائية بسيطة. وتوجد نظرية مشابهة للنظرية (٨-٢) تعبّر عن $V(\bar{y}_{sy})$ بدلالة التباين الموافق لعيّنة عشوائية طبقية تتألف فيها الطبقات من الوحدات الـ k الأولى، الوحدات الـ k التي تليها، وهكذا. وفي رموزنا يشير الدليل j في y_{ij} إلى الطبقة. وسنكتب متوسط الطبقة على الشكل $\bar{y}_{.j}$.

نظرية (٨ - ٣)

$$V(\bar{y}_{sy}) = \frac{S^2_{wst}}{n}\left(\frac{N-n}{N}\right)[1 + (n-1)\rho_{wst}] \quad (8.11)$$

حيث ،

$$S^2_{wst} = \frac{1}{n(k-1)} \sum_{j=1}^{n} \sum_{i=1}^{k} (y_{ij} - \bar{y}_{.j})^2 \quad (8.12)$$

وهو التباين بين الوحدات الواقعة في الطبقة نفسها. ونستخدم المقام $n(k\text{-}1)$ لأن كلاً من الطبقات الـ n تسهم بـ $(k\text{-}1)$ درجة من الحرية. وبالإضافة إلى ذلك فإن ،

$$\rho_{wst} = \frac{E(y_{ij} - \bar{y}_{.j})(y_{iu} - \bar{y}_{.u})}{E(y_{ij} - \bar{y}_{.j})^2} \quad (8.13)$$

وهذه الكمية هي معامل الارتباط بين انحرافات أزواج من المفردات، موجودة ضمن العيّنة النمطية نفسها، عن متوسطات الطبقات.

$$\rho_{wst} = \frac{2}{n(n-1)(k-1)} \sum_{i=1}^{k} \sum_{j<u} \frac{(y_{ij} - \bar{y}_{.j})(y_{iu} - \bar{y}_{.u})}{S_{wst}^2} \qquad (8.14)$$

والبرهان مشابه لما رأيناه في النظرية (٨-٢).

نتيجة

للعيّنة النمطية نفس دقة العيّنة العشوائية الطبقية الموافقة، بوحدة واحدة ضمن كل طبقـة، إذا كان $\rho_{wst} = 0$. وهذا ناتج عن كون $V(\bar{y}_{st})$ في هذا النوع من العيّنات العشوائية الطبقية (نظرية ٥-٣، نتيجة ٣) مساوٍ لـ

$$V(\bar{y}_{st}) = \left(\frac{N-n}{N}\right)\frac{S_{wst}^2}{n} \qquad (8.15)$$

وهنـاك علاقـات أخرى حول $V(\bar{y}_{sy})$، مناسبة لمجتمع ذاتي الترابط، طوّرها W.G. madow و L.H. Madow (1944) اللذين قاما بالـدراسـة النظرية الأولى لدقة المعاينة النمطية.

مثـال

يتعلق البيان الإحصائي في الجدول (٨ ـ ٣) بمجتمع اصطناعي صغير يُفصح عن اتجـاه صاعد بثبات تقريبًا. لدينا $n=4, k=10, N=40$. ويمثل كل عمود عيّنة نمطية، والصفوف هي الطبقات. ويوضح المثال الحالة التي يكون فيها الارتباط ضمن العيّنة إيجابيًا. فعلى سبيل المثال، يقع كل من الأعداد الأربعة 18, 6, 0 و 26 في العيّنة الأولى تحت متوسط الطبقة التي ينتمي إليها العدد. ويبقى هذا صحيحًا في العيّنات النمطية الخمس الأولى، مع قليل من الاستثناءات وفي العيّنات الخمس الأخيرة يبقى الانحراف عن متوسط الطبقة إيجابيًا في معظم الحالات. وهكذا تكون الحدود الجدائية في ρ_{wst} موجبة في معظمها. ومن النظرية (٨ ـ ٣) نتوقع أن تكون المعاينة النمطية أقل دقة من المعاينة العشوائية الطبقية مع وحدة واحدة من كل طبقة.

ويُحسب التباين $V(\bar{y}_{sy})$ مباشرة من مجاميع العيّنات النمطية على الشكل،

$$V(\bar{y}_{sy}) = V_{sy} = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.} - \bar{Y})^2 = \frac{1}{n^2 k}\sum_{i=1}^{k}(n\bar{y}_{i.} - n\bar{Y})^2$$

$$= \frac{1}{160}\left[(50)^2 + (58)^2 + \cdots + (88)^2 - \frac{(727)^2}{10}\right] = 11.63$$

جدول (٨ ـ ٣) بيان إحصائي لـ 10 عيّنات نمطية حيث $N=kn=40, n=4$

الطبقة	1	2	3	4	5	6	7	8	9	10	متوسط الطبقة
	الأرقام المسلسلة للعيّنات النمطية										
I	0	1	1	2	5	4	7	7	8	6	4.1
II	6	8	9	10	13	12	15	16	16	17	12.2
III	18	19	20	20	24	23	25	28	29	27	23.3
IV	26	30	31	31	33	32	35	37	38	38	33.1
المجاميع	50	58	61	63	75	71	82	88	91	88	72.7

وفي معاينة عشوائية بسيطة أو معاينة طبقية نحتاج إلى تحليل تباين المجتمع إلى «ما بين الصفوف» و«ما ضمن الصفوف» . وهذا مبيّن في الجدول (٨ ـ ٤) . ومنه تكون تباينات تقديرات المتوسطات مستخدمين عيّنات عشوائية بسيطة وعيّنات عشوائية طبقية كما يلي :

جدول (٨ ـ ٤) تحليل التباين

	درجات الحرية	مجموع المربعات	متوسط المربعات
ما بين الصفوف (الطبقات)	3	4828.3	
ما ضمن الطبقات	36	485.5	$13.49 = S^2_{wst}$
المجموع	39	5313.8	$136.25 = S^2$

$$V_{ran} = \left(\frac{N-n}{N}\right)\frac{S^2}{n} = \frac{9}{10}\cdot\frac{136.25}{4} = 30.66$$

$$V_{st} = \left(\frac{N-n}{N}\right)\frac{S^2_{wst}}{n} = \frac{9}{10}\cdot\frac{13.49}{4} = 3.04$$

وكلتا المعاينتين العشوائية الطبقية والنمطية أكثر فعالية بكثير من المعاينة العشوائية البسيطة، ولكن المعاينة النمطية، كما هو منتظر، أقل دقة من المعاينة العشوائية الطبقية.

ويبيّن الجدول (٨ ـ ٥) المعلومات الإحصائية نفسها، مع عكس ترتيب الملاحظات في الطبقتين الثانية والرابعة. وتأثير ذلك هو جعل ρ_{wst} سالبًا، لأنه يجعل معظم الحدود الجدائية بين الانحرافات عن متوسطات الطبقات سالبة وذلك من أجل أزواج من الملاحظات واقعة في العيّنة النمطية نفسها. وفي العيّنة النمطية الأولى، مثلاً، تصبح الانحرافات عن متوسطات الطبقات الآن كما يلي: 4.9,-5.3,4.8,-4.1- . ومن بين الجداءات الستة لأزواج الانحرافات نجد أن أربعة منها سالبة. وعلى وجه التقريب تنطبق الحالة نفسها في كل عيّنة نمطية.

جدول (٨ ـ ٥) البيان الإحصائي في الجدول (٨ـ٣) مع عكس الترتيب في الطبقتين II و IV

الطبقة	أرقام التسلسل للعيّنة النمطية 1	2	3	4	5	6	7	8	9	10	متوسط الطبقة
I	0	1	1	2	5	4	7	7	8	6	4.1
II	17	16	16	15	12	13	10	9	8	6	12.2
III	18	19	20	20	24	23	25	28	29	27	23.3
IV	38	38	37	35	32	33	31	31	30	26	33.1
المجموع	73	74	74	72	73	73	73	75	75	65	72.7

وهذا التغير لا يؤثر في V_{ran} و V_{st}. وتؤدي في حالة المعاينة النمطية إلى زيادة مثيرة في الدقة، كما سنرى عند مقارنة مجاميع العيّنات النمطية في كل من الجدولين (٨ ـ ٥) و(٨ ـ ٣). ولدينا الآن،

$$V_{sy} = \frac{1}{160}\left[(73)^2 + (74)^2 + \cdots + (65)^2 - \frac{(727)^2}{10}\right] = 0.46$$

ومن الممكن أحيانًا استغلال هذه النتيجة بترقيم الوحدات بحيث نُحدث ارتباطًا سلبيًا ضمن الطبقات، ومن المطلوب توافر معرفة دقيقة بالاتجاهات ضمن المجتمع. وعلى أي حال، وكما سنرى فيما بعد، فإن الحالة في الجدول (٨ ـ ٥) هي حالة يصعُب معها الحصول من العيّنة على تقدير جيد للخطأ المعياري لِـ $\bar{y}_{sy}$.

(٨ ـ ٤) مقارنة المعاينة العشوائية الطبقية بالمعاينة النمطية

إن ما تنجزه المعاينة النمطية بالمقارنة مع المعاينة العشوائية الطبقية أو البسيطة يعتمد إلى حد كبير على خواص المجتمع. وتوجد مجتمعات تكون المعاينة النمطية فيها دقيقة للغاية، وأخرى تكون فيها أقل دقة من المعاينة العشوائية البسيطة. ومع بعض المجتمعات وبعض قيم n يمكن لِـ $V(\bar{y}_{sy})$ حتى أن يزداد عند أخذ عيّنة أكبر، وهو ابتعاد مدهش عن السلوك الجيّد. وهكذا فإنه من الصعب إعطاء نصيحة عامة حول الحـالات التي يوصى فيها بالمعاينة النمطية. ومعرفة بنية المجتمع ضرورية من أجل الاستخدام الأكثر فعالية لها.

وقد اتُّبع خطان من البحث في هذه المسألة. أحدهما هو مقارنة أنواع مختلفة من المعـاينة في مجتمعات اصطناعية يكون فيها y_i دالة بسيطة في i. والآخر هو أن نقوم بالمقارنة في حالة مجتمعات واقعية. ونقدم بعض النتائج الرئيسة في الفقرات القادمة.

(٨ ـ ٥) مجتمعات ذات ترتيب «عشوائي»

تُستخدم المعاينة النمطية أحيانًا، لسهولتها، في مجتمعات يكون ترقيم الوحدات فيها عشوائيًا فعلاً. ويكون الأمر كذلك عند معاينة مجموعة بطاقات مرتبة أبجديًا وفقًا لأسماء الكنية، إذا لم يكن للمفردة التي نقيسها أية علاقة بكنية الشخص. وسوف لا يوجد عندئذ أي اتجاه أو تقسيم إلى طبقات في y_i ونحن نمضي على طول هذه البطاقات، كما لا يوجد أي ارتباط بين القيم المتجاورة.

وفي هذه الحالة، يمكن أن نتوقع نوعًا من التكافؤ بين المعاينة النمطية والمعاينة العشوائية البسيطة، وأن يكون لهما التباين نفسه. وليس هذا صحيحًا بالضبط في أي مجتمع منته بمفرده مع قيم معطاة لِـ n و k، باعتبار أن V_{sy}، المبني على k من درجات الحرية فقط، سيكون غريب الأطوار عندما يكون k صغيراً، ويمكن أن يتمخض عن قيمة أكبر أو أصغر من V_{ran}. وهناك نتيجتان تبينان أن التباينين هما، في المتوسط، متساويان.

نظرية (٨ ـ ٤)

لنعتبر كل الـ !N من المجتمعات المشكلة من التباديل الـ !N لأي مجموعة من الأعداد $y_1, y_2, \ldots, y_N$ وبأخذ المتوسط فوق هذه المجتمعات المنتهية نجد:

$$E(V_{sy}) = V_{ran} \tag{8.16}$$

ونلاحظ أن V_{ran} يبقى نفسه في جميع التباديل.

وهذه النتيجة التي برهنها W.G. Madow و L.H. Madow (1944)، تبيّن أنه إذا أمكن اعتبار ترتيب المفردات في مجتمع معيّن منته وكأنه مسحوب عشوائيًا من التباديل الـ !N ، فعندئذ تكون المعاينة النمطية مكافئة، في المتوسط، للمعاينة العشوائية البسيطة.

والطريقة الثانية هي أن نعتبر المجتمع المنتهي وكأنه مسحوب عشوائيًا من مجتمع لانهائي فوقي له خواص معينة. والخاصة التي بُرهنت لا تنطبق على أي مجتمع منته بمفرده (ونعني على أي مجموعة محددة من القيم y_1 و $y_2, \ldots, y_N$) ولكنها تنطبق على متوسط كل المجتمعات المنتهية التي يمكن سحبها من المجتمع اللانهائي.

ويرمز E لعملية أخذ المتوسط فوق جميع المجتمعات المنتهية التي يمكن سحبها من هذا المجتمع الفوقي.

نظرية (٨ ـ ٥)

إذا كانت المتغيرات $y_i (i=1,2,\ldots,N)$ مسحوبة عشوائيًا من مجتمع فوقي فيه:

$$\mathscr{E}y_i = \mu, \qquad \mathscr{E}(y_i - \mu)(y_j - \mu) = 0 \quad (i \neq j), \qquad \mathscr{E}(y_i - \mu)^2 = \sigma_i^2$$

فعندئذ،

$$\mathscr{E}V_{sy} = \mathscr{E}V_{ran}$$

والشروط الحاسمة هي أن يكون لجميع المقادير y_i المتوسط μ نفسه أي عدم وجود أي اتجاه، وألا يوجد ارتباط خطي بين القيمتين y_i و y_j في نقطتين مختلفتين. ويمكن أن يتغير التباين σ_i^2 من نقطة إلى أخرى في السلسلة.

برهان

في أي مجتمع معيّن منته لدينا:

$$V_{ran} = \frac{N-n}{Nn} \frac{\sum_{i=1}^{N} (y_i - \bar{Y})^2}{N-1}$$

والآن

$$\sum_{i=1}^{N} (y_i - \bar{Y})^2 = \sum_{i=1}^{N} [(y_i - \mu) - (\bar{Y} - \mu)]^2$$

$$= \sum_{i=1}^{N} (y_i - \mu)^2 - N(\bar{Y} - \mu)^2$$

وبما أن y_i و $y_j (i \neq j)$ غير مرتبطين، فلدينا:

$$\mathscr{E}(\bar{Y} - \mu)^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sigma_i^2 \tag{8.17}$$

ومنه،

$$\mathscr{E} V_{ran} = \frac{N-n}{Nn(N-1)} \left(\sum_{i=1}^{N} \sigma_i^2 - N \frac{\sum \sigma_i^2}{N^2} \right) \tag{8.18}$$

وهذا يعطي،

$$\mathscr{E} V_{ran} = \frac{N-n}{N^2 n} \sum_{i=1}^{N} \sigma_i^2 \tag{8.19}$$

لنلتفت إلى V_{sy}، ولنرمز بـ $\bar{y}_u$ لمتوسط العيّنة النمطية الـ u. ففي أي مجتمع منته معين لدينا،

$$V_{sy} = \frac{1}{k} \sum_{u=1}^{k} (\bar{y}_u - \bar{Y})^2 \tag{8.20}$$

$$= \frac{1}{k} \left[\sum_{u=1}^{k} (\bar{y}_u - \mu)^2 - k(\bar{Y} - \mu)^2 \right] \tag{8.21}$$

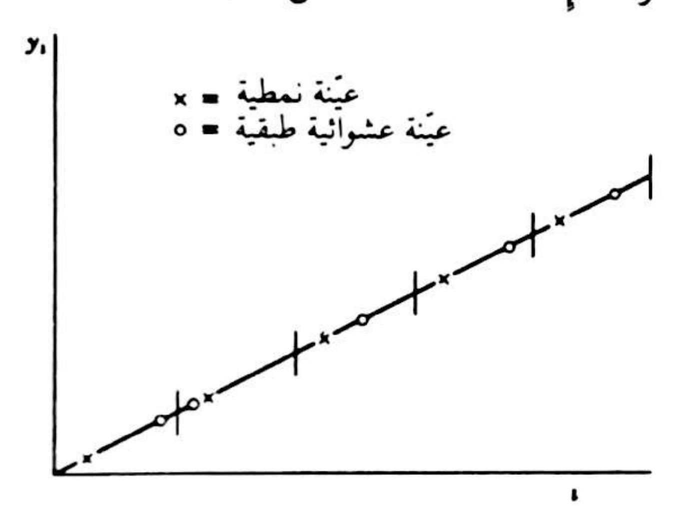
وبالاستناد إلى نظرية تباين متوسط عيّنة غير مرتبطة من مجتمع لا نهائي نجد،

$$\mathscr{E} V_{sy} = \frac{1}{k} \left(\frac{\sum_{i=1}^{N} \sigma_i^2}{n^2} - \frac{k \sum_{i=1}^{N} \sigma_i^2}{N^2} \right) \tag{8.22}$$

$$= \frac{N-n}{N^2 n} \sum_{i=1}^{N} \sigma_i^2 = \mathscr{E} V_{ran} \tag{8.23}$$

(٨ - ٦) مجتمعات ذات اتجاه خطي

إذا تألف المجتمع من اتجاه خطي فقط، كما هو موضح في الشكل (٨ - ٢)، فمن السهل، إلى حد ما، أن نخمّن طبيعة النتائج . ويبدو من الشكل (٨ - ٢) كما لو أن كلاً من V_{sy} و V_{st} (بوحدة واحدة من كل طبقة) سيكون أصغر من V_{ran} . وفضلاً عن ذلك سيكون V_{sy} أكبر من V_{st} ذلك لأنه إذا كانت العيّنة النمطية منخفضة جدًّا في إحدى الطبقات فإنها ستكون منخفضة جدًّا أيضًا في جميع الطبقات، بينما تقدم المعاينة العشوائية الطبقية فرصة لإلغاء أخطاء ما ضمن الطبقة.



شكل (٨ - ٢) المعاينة النمطية من مجتمع ذي اتجاه خطي

ولدراسة التأثيرات رياضيًا يمكن أن نفترض أن $y_i=i$ وعندئذ لدينا:

$$\sum_{i=1}^{N} i = \frac{N(N+1)}{2}, \qquad \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}$$

وتباين المجتمع S^2 معطى بالعلاقة،

$$S^2 = \frac{1}{N-1}(\sum y_i^2 - N\bar{Y}^2)$$

$$= \frac{1}{N-1}\left[\frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{4}\right] = \frac{N(N+1)}{12} \qquad (8.24)$$

وهكذا فإن تباين متوسط عينة عشوائية بسيطة هو،

$$V_{ran} = \frac{N-n}{N} \cdot \frac{S^2}{n} = \frac{n(k-1)}{N} \cdot \frac{N(N+1)}{12n} = \frac{(k-1)(N+1)}{12} \qquad (8.25)$$

ولإيجاد التباين ضمن الطبقات S_w^2 ، نحتاج فقط لوضع k بدلاً من N في (8.24) وهذا يعطي :

$$V_{st} = \frac{N-n}{N} \cdot \frac{S_w{}^2}{n} = \frac{n(k-1)}{nk} \cdot \frac{k(k+1)}{12n} = \frac{(k^2-1)}{12n} \tag{8.26}$$

وفي المعاينة النمطية يتجاوز متوسط العينة الثانية متوسط العيّنة الأولى بمقدار 1 ، بينما يتجاوز متوسط الثالثة ذلك المتعلق بالثانية بمقدار 1 ، وهلمّ جرًّا. وهكذا فإنه يمكن استبدال الأعداد $1,2,\ldots,k$ بالمتوسطات $\bar{y}_u$. ومنه، وبتطبيق (8.24) من جديد، نجد :

$$\sum_{u=1}^{k} (\bar{y}_u - \bar{Y})^2 = \frac{k(k^2-1)}{12}$$

وهذا يعطي ،

$$V_{sy} = \frac{1}{k} \sum (\bar{y}_u - \bar{Y})^2 = \frac{k^2-1}{12} \tag{8.27}$$

ونستنتج من العلاقات (8.25) ، (8.26) ، (8.27) كما هو منتظر :

$$V_{st} = \frac{k^2-1}{12n} \leq V_{sy} = \frac{k^2-1}{12} \leq V_{ran} = \frac{(k-1)(N+1)}{12} \tag{8.28}$$

وتحصل المساواة فقط عندما $n=1$ ، وسواء أكنّا نظن بوجود اتجاه خطي أم لا ، فإن العيّنة النمطية أكثر فعالية بكثير من العيّنة العشوائية الطبقية .

(٨ ـ ٧) طرق لمجتمعات ذات اتجاهات خطّية

في حال وجود اتجاه خطي يمكن تحسين أداء المعاينة النمطية بعدة طرق . إحداها هي استخدام عيّنة متموضعة مركزيًا ، وأخرى هي تحويل التقدير من متوسط غير مرجّح إلى متوسط مرجّح ، تتلقى فيه كل العناصر الداخلية في العيّنة الترجيحة 1 (قبل القسمة على n)، ولكن العنصرين الأول والأخير يتلقيان ترجيحتين مختلفتين عن ذلك . وإذا كان العدد العشوائي المسحوب بين 1 و k مساويًا لـ i فتكون هاتان الترجيحتان ،

$$1 \pm \frac{n(2i-k-1)}{2(n-1)k} \tag{8.29}$$

وتُستخدم إشارة + للعنصر الأول وإشارة - للعنصر الأخير ومن أجل أي قيمة لِـ i يكون مجموع الترجيحتين، بوضوح، مساويًا لِـ 2 . ويمكن للقارىء أن يتحقق من أنه إذا انطوى المجتمع على اتجاه خطي وكان $N=nk$ فإن متوسط العيّنة المرجّح يعطي المتوسط الصحيح للمجتمع. وقد درس Yates (1946) ما ينجزه تصحيحا النهاية هذين، وإليه يعود هذان التصحيحان. وقد وسّع Bellhouse و Rao (1975) تصحيحي Yates إلى الحالة $N=nk$ وذلك عندما تُسحب العيّنة النمطية وفقًا لطريقة Lahiri الدائرية (فقرة ٨-١)، التي تضمن ثبات n . وكما سبق تُطبّق الترجيحات المختلفة عن الواحد على رقمي العيّنة الأول والأخير معبرًا عنهما وفق التسلسل الرقمي الأصلي للمجتمع. وعلى سبيل المثال، إذا كان عدد البداية العشوائي عند سحب العيّنة هو 19 ، حيث $N=23$ ، $n=5$ فوحدات العيّنة هي الوحدات 19 ,1 ,6 ,11 ,16 ، والعنصران الأول والأخير هما y_1 و y_{19} وتبرز حالتان:

حالة ١ : i صغير بحيث إن $i=(n-1)k\leq N$ ، وعندئذ نحصل على n من الوحدات دون تجاوز y_N . وتكون الترجيحتان للعنصر الأول (+) وللعنصر الأخير (−) هما:

$$1\pm\frac{n[2i+(n-1)k-(N+1)]}{2(n-1)k} \tag{8.30}$$

حالة ٢ : $i=(n-1)k>N$. ليكن a_2 عدد وحدات العيّنة التي نحصل عليها بعد تجاوز y_N . وهكذا يكون $n_2=4$ في حالة $i=19$. فتكون الترجيحتان للعنصر الأول (+) وللعنصر الأخير (−) هما:

$$1\pm\frac{n}{2(N-k)}\left[2i+(n-1)k-(N+1)-2n_2\frac{N}{n}\right] \tag{8.31}$$

وفي كلتا الحالتين تتلقى العناصر الداخلية في العيّنة ترجيحة 1 في مجموع العيّنة. وفي حالة $N=23$, $n=k=5$, $i=19$, $n_2=4$ تكون الترجيحتان الأولى والأخيرة هما $1\pm(-7/18)$. وبالتالي يتلقى y_1 ترجيحة 11/18 بينما يتلقى y_{19} ترجيحة 25/18 .

وهناك طريقتان بديلتان تحاولان تغيير طريقة اختيار العيّنة بحيث لا يتأثر متوسط العيّنة باتجاه خطي. ومع $N=nk$ و n زوجي، يقترح Sethi (1965) طريقة نقسم بموجبها

المجتمع إلى $n/2$ طبقة حجم كل منها $2k$ ، ونختار وحدتين متساويتي البعد عن نهايتي كل طبقة. ومع عدد بداية عشوائي i يكون الـ $n/2$ زوجًا من الوحدات هي تلك التي أرقامها

$$[i+2jk, 2(j+1)k-i+1], \qquad j=0, 1, 2, \ldots \tfrac{1}{2}n-1 \qquad (8.32)$$

ويحـذف هذا الاختيار تأثـير وجـود اتجـاه خطي في أي طبقـة تتضمن k من الـوحـدات، حتى ولـو تغيّر الميل الخـطي من طبقـة إلى طبقـة. وقـد دعـا Murthy (1967) الطريقة بطريقة المعاينة النمطية المتوازنة.

وتختار الطريقة المعدّلة لِـ Singh وآخرين (1968) أزواجًا من الوحدات المتساوية البعـد عن نهايتي المجتمـع. ومع n زوجي، تكون الأزواج متساوية البعد التي تبدأ بالوحدة i $(i=1,2,\ldots,k)$ ، وعددها $\frac{n}{2}$ هي،

$$[i+jk, (N-jk)-i+1], \qquad j=0, 1, 2, \ldots \tfrac{1}{2}n-1 \qquad (8.33)$$

وفي حالـة n فردي يمضي الـدليل j ، في هاتـين الـطريقتـين، إلى $\frac{1}{2}(n-1)-1$ في (8.32) و (8.33) . وتضيف الـطريقة المتوازنة (8.32) عنصر العيّنة الباقي الواقع قرب النهاية عند $[i+(n-1)k]$ ؛ بينما تضيف الطريقة المعدّلة العنصر $[i+\frac{1}{2}(n-1)k]$ الواقع قرب الوسط. ولا يُحذف تأثير الاتجاه الخطي في $\bar{y}$ تمامًا في حالة n فردي.

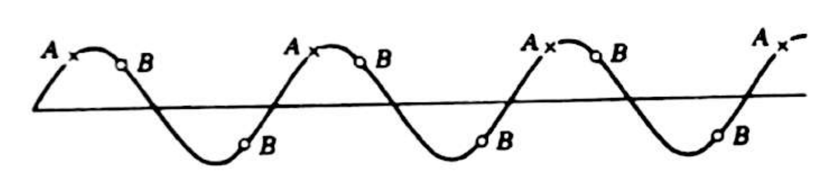
وقد أُجريت مقارنات إنجاز هاتين الطريقتين مع تصحيحات Yates ومع المعاينة النمـطية العـادية في نماذج مجتمعـات فوقية تمثـل اتجـاهـات خطية وقـطع ـ مكافئية (شلجمية) وتغيرات دورية وذاتية الارتباط (Bellhouse و Rao, 1975) وذلك في قليل من المجتمعـات الواقعية الصغيرة، من قِبل هذين الباحثين ومن قِبل Singh (اتصال شخصي). وبصورة عامة، كانت الطرق الثلاث (Yates المتوازنة، المعدّلة) متماثلة في أدائها، ومتفوقة على المعاينة النمطية العادية في حال وجود اتجاه خطي أو قطع ـ مكافئي (شـلجمي).

والمجتمع المذكور في الجدول (٨ ـ ٣)، على سبيل المثال، هو مجتمع ينبغي أن تنجـز هذه الـطرق فيه إنجـازًا حسنًـا جدًّا. وقـد أعـطت المعـاينة النمطية العادية

$V_{sy}=11.63$. والتبـاينـات المقـارنـة في الطرق الأخرى ($k=10, n=4$) هي : Yates ، 1.29 ؛ Sethi (متوازنة)، 0.46 ؛ Singh (معدّلة)، 0.34 . ويتفق أن تكون الطريقة المتوازنة هي تلك التي نحصل عليها من الجدول (٨ ـ ٥) عندما تتبادل الطبقتان II و IV موقعيهما في الجدول (٨ ـ ٧).

(٨ ـ ٨) مجتمعات ذات تغير دوري

إذا كان المجتمع ينطوي على اتجاه دوري، مثلاً منحنى جيبي بسيط، فتعتمد فعـالية العيّنـة النمطية عندئذ على قيمة k . ويمكن رؤية ذلك تصويريًا في الشكل (٨-٣). وفي هذا التمثيل تكـون المـلاحـظة y_i هي ارتفـاع المنحني. وتمثـل نقـاط العيّنـة A الحالة الأقل تلاؤمًا مع المعاينة النمطية، وتصحّ هذه الحالة حيثما يكون k مساويًا لدور المنحني الجيبي أو لأحد مضاعفاته. وكل الملاحظات ضمن العينة النمطية تبقى نفسها تمامًا، بحيث لا تزيد دقة العيّنة على ملاحظة واحدة مأخوذة من المجتمع عشوائيًا.



شكل (٨ ـ ٣) تغير دوري

وتقع الحالة الأكثر تلاؤمًا (العيّنة B) عندما يكون k مساويًا إلى عدد فردي من أنصاف الدور، ولكل عينة نمطية متوسط يساوي المتوسط الحقيقي للمجتمع تمامًا، باعتبار أن الانحرافات المتتالية فوق وتحت الخط الأوسط تلغي بعضها بعضًا. ولذلك فإن تباين المعاينة بالنسبة للمتوسط يساوي الصفر. وللعيّنة بين هاتين الحالتين درجات مختلفة من الفعالية، وذلك يتوقف على العلاقة بين k وطول الموجة.

ومن غير المرجّح مواجهة مجتمعات تُفصح عن منحنى جيبي تمامًا في التطبيق العملي. إلا أن المجتمعات ذات الاتجاه الدوري الواضح إلى حد ما، ليست أمرًا نادر

الحدوث. وكأمثلة نسوق تدفق حركة السير، بعد نقطة على طريق، فوق الساعات الأربع والعشرين من اليوم، ومبيعات مخزن فوق الأيام السبعة من الأسبوع. ولتقدير متوسط فوق فترة دورية نجد بوضوح أنه ليس من الحكمة أن نأخذ عيّنة نمطية عند الساعة الرابعة بعد الظهر يوميًا أو كل يوم ثلاثاء. وبدلاً من ذلك نقول إن الاستراتيج الصحيح هو أن تتهادى العيّنة فوق المنحنى الدوري، بأن نرى مثلاً في حالة مبيعات مخزن أن كل يوم من أيام الأسبوع ممثل على قدم المساواة.

ولبعض المجتمعات نوع من التأثير الدوري الأقل وضوحًا. ففي سلسلة من قوائم الأجور الأسبوعية، في قطاع صغير من مصنع، يمكن أن يردَ العمال دائمًا وفق ترتيب ما، وقد تحوي بين 19 و 32 اسمًا كل أسبوع. وعيّنة نمطية من 1 من عشرين اسمًا فوق فترة عدة أسابيع، يمكن أن تتألف بصورة رئيسة من سجلات عامل أو سجلات عاملين أو ثلاثة عمال. وبصورة مماثلة يمكن أن تحوي عيّنة نمطية من الأسماء من دليل مدينة العديد من أرباب الأسر، أو العديد من الأطفال. وإذا كان هناك وقت لدراسة البنية الدورية فيمكن عادة تصميم العيّنة النمطية بحيث نستفيد من هذه الدراسة. والفشل في ذلك، مع أننا نشك بوجود تأثير دوري غير معروف جيدًا، يجعل العينة العشوائية البسيطة أو الطبقية مفضلة على العيّنة العشوائية النمطية.

وفي بعض المجتمعات الواقعية قد يوجد تغير شبه دوري يصعُب توقعه. وقد وجد L.H. Madow (1946) بيّنة تشير إلى هذا في مشتل من الشجيرات الصغيرة يشكل مجتمعًا صغيرًا إلى حد ما ($N=420$). وقد ناقش Finney (1950) ظاهرة مشابهة بالنسبة لحجم الأخشاب للشريط الواحد في غابة Dehra Dun، مع أن Milne (1959) اقترح عند إعادة دراسة البيان الإحصائي أن الدورية الظاهرة قد تكون نتاجًا لعملية القياس. وتأثير شبه الدورية أن المعاينة النمطية تؤدي أداءً فقيرًا من أجل بعض قيم n وأداءً واضح الجودة من أجل قيم أخرى. ومن غير المعروف ما إذا كان هذا التأثير يحدث بصورة متكررة. ويذكر Matern (1960) أمثلة يمكن أن تُنتج فيها قوى طبيعية (المدّ والجزر مثلًا) تغيرًا مكانيًا دوريًا، إلا أنه من أصحاب الرأي القائل بأنه لم يُعثر على حالة واضحة في المسوح الإحصائية للغابات.

(٨ - ٩) المجتمعات ذاتية الارتباط

في العديد من المجتمعات الواقعية، يوجد سبب للتوقع بأن ملاحظتين y_i ، y_j ستكونان أميل إلى التشابه، عندما تكون i و j قريبتين من بعضهما في السلسلة، مما لو كانتا بعيدتين عن بعضهما. ويحدث هذا حيثما تُنتج القوى الطبيعية تغيراً بطيئاً ونحن نمضي على طول السلسلة. وفي نموذج رياضي لهذا التأثير يمكن أن نفترض y_i و y_j مرتبطين إيجابيًا، والارتباط بينهما تابع فقط للمسافة التي تفصلهما $(i-j)$ ، ويتناقص كلما تزايدت هذه المسافة. ومع أن هذا النموذج مبسط أكثر من اللازم إلا أنه يمكن أن يمثل إحدى المقوّمات البارزة للعديد من المجتمعات الواقعية.

ولكي نتقصّى ما إذا كان هذا النموذج ينطبق على مجتمع ما أم لا، يمكن حساب مجموعة الارتباطات ρ_u في أزواج من المفردات تبعد عن بعضهما بمقدار u من الوحدات، ونرسم هذا الارتباط بيانيًا في مقابل u . ويسمى هذا المنحني أو الدّالة التي تمثله «مصوّر الارتباط». وحتى إذا كان النموذج مشروعًا، فسوف لا يكون مصور الارتباط دالّة ممهّدة في أي مجتمع منته، وذلك بسبب عدم الانتظام الذي تسببه الطبيعة المنتهية للمجتمع. وفي أي مقارنة للمعاينتين النمطية والعشوائية الطبقية، تحت هذا النموذج، فإن عدم الانتظام هذا يجعل اشتقاق النتائج، لأي مجتمع منته بمفرده، أمرًا صعبًا. ويمكن القيام بالمقارنة فوق متوسط سلسلة كاملة من المجتمعات المنتهية، المسحوبة عشوائيًا من مجتمع فوقي لا نهائي ينطبق عليه النموذج. وقد طبقنا هذه الطريقة سابقًا في النظرية (٨ - ٥) وفي الفقرتين (٦ - ٧) و(٧ - ٨).

وهكذا فإننا نفترض أن الملاحظات $y_i(i=1,2,\dots,N)$ مسحوبة من مجتمع فوقي فيه:

$$\mathscr{E}(y_i)=\mu, \quad \mathscr{E}(y_i-\mu)^2=\sigma^2, \quad \mathscr{E}(y_i-\mu)(y_{i+u}-\mu)=\rho_u\sigma^2 \qquad (8.34)$$

حيث، $\rho_u \geq \rho_v \geq 0$ حيثما يكون $u<v$.

وسحب مجموعة واحدة من قيم y_i من هذا المجتمع الفوقي يخلق مجتمعًا منتهيًا واحدًا حجمه N .

ونرمز لمتوسط التباين الموافق لمعاينة نمطية بـ

$$\mathscr{E}V_{sy} = \mathscr{E}E(\bar{y}_{sy} - \bar{Y})^2$$

ومن السهـل أن نبيِّن، في هذا الصف من المجتمعات، أن المعاينة العشوائية الطبقية متفوقة على المعاينة العشوائية البسيطة، إلا أنه لا يمكن إرساء نتيجة عامة حول المعاينة النمطية. وتوجد ضمن الصف مجتمعات فوقية تتفوق المعاينة النمطية فيها على المعاينة العشوائية الطبقية، ولكن توجد أيضًا مجتمعات فوقية تكون فيها المعاينة النمطية متخلفة حتى عن المعاينة العشوائية البسيطة، وذلك من أجل قيم معينة لِـ k.

ويمكن الحصـول على نظرية عامـة إذا فرضنـا بالإضافة إلى ذلك أن مصوَّر الارتباط مقعَّر إلى الأعلى.

نظرية (٨ ـ ٦)

إذا كان لدينا بالإضافة إلى الشرط (8.34)

$$\delta_u^2 = \rho_{u+1} + \rho_{u-1} - 2\rho_u \geq 0 \quad [u = 2, 3, \ldots, (kn-2)] \tag{8.35}$$

فعندئذ،

$$\mathscr{E}V_{sy} \leq \mathscr{E}V_{st} \leq \mathscr{E}V_{ran} \tag{8.36}$$

وذلك من أجل أي حجم للعيِّنة، مالم يكن $\delta_u^2 = 0$، $u=2,3,\ldots,(kn-2)$ فلدينا فضلاً عن ذلك:

$$\mathscr{E}V_{sy} < \mathscr{E}V_{st} \tag{8.37}$$

وقـد قدّم Cochran (1946) برهـانًـا. ويوضح عرض لمخطط البرهان في حالة $n=2$ الدور الذي يلعبه شرط «التقعر إلى الأعلى». وفي العيِّنة النمطية يكون عنصرا الزوج دائمًا على مسافة k من الوحدات عن بعضهما. وبالتالي،

$$\mathscr{E}V(\bar{y}_{sy}) = \tfrac{1}{4}(\sigma^2 + \sigma^2 + 2\rho_k\sigma^2) = \tfrac{1}{2}\sigma^2(1 + \rho_k) \tag{8.38}$$

وفي حالة عيِّنة طبقية توجد k من المواقع الممكنة للوحدة المسحوبة من كل طبقة مما يؤدي إلى k^2 من تراكيب المـواقـع. وأعـداد التراكيب التي تبعد عن بعضها بمقدار 1 ، 2 ، . . . ، $(2k-1)$ من الوحدات هي كما يلي:

المسافة	$1 \quad 2\ldots(k-1) \quad k \quad (k+1)\ldots(2k-1)$	المجموع
العدد	$1 \quad 2\ldots(k-1) \quad k \quad (k-1)\ldots \quad 1$	k^2

وبالتالي يمكن كتابة القيمة المتوسطة لِـ $V(\bar{y}_{st})$ مأخوذة فوق الـ k^2 تركيبًا كما يلي:

$$\mathscr{E}V(\bar{y}_{st}) = \frac{\sigma^2}{2k^2}\left[\sum_{u=1}^{k-1} u(2+\rho_u+\rho_{2k-u}) + k(1+\rho_k)\right] \qquad (8.39)$$

وبصورة مماثلة يمكن التعبير عن $\mathscr{E}V(\bar{y}_{st})$ على الشكل،

$$\mathscr{E}V(\bar{y}_{sy}) = \frac{\sigma^2}{2k^2}\left[\sum_{u=1}^{k-1} u(2+2\rho_k) + k(1+\rho_k)\right] \qquad (8.40)$$

ومنه،

$$\mathscr{E}V(\bar{y}_{st}) - \mathscr{E}V(\bar{y}_{sy}) = \frac{\sigma^2}{2k^2}\left[\sum_{u}^{k-1} u(\rho_u+\rho_{2k-u}-2\rho_k)\right] \qquad (8.41)$$

ولكن إذا كان،

$$\rho_{u+1}+\rho_{u-1} \geq 2\rho_u \qquad (u=2, 3, \ldots)$$

فمن السهـل تبيان أن كل حد ضمن القـوسـين موجب. وهكـذا يكتمل البرهان. وبـاختصـار، فمتوسط المسافة هو k في كل من العيّنتين النمطية والطبقية، إلا إنه، وبسبب التقعر، تخسر العيّنة الطبقية من دقتها عندما تقلّ المسافة عن k أكثر مما تكسبه عندما تتجاوز المسافة k .

وبـرهن Quenouille (1949) أن المتبـاينـات في النظرية (٨-٦) تبقى صحيحة عندما نستغني عن شرطين من الشروط بحيث إن،

$$\mathscr{E}(y_i) = \mu_i, \qquad \mathscr{E}(y_i-\mu_i)^2 = \sigma_i^2 \qquad (8.42)$$

وفي هذه الحالة يزداد كل من التباينات المتوسطة الثلاثة بالمقدار نفسه.

وإلى الحد الذي يتعلق بالتطبيقات العملية فإن مصوّرات الارتباط المقعرة إلى الأعلى كانت قد اقتُرحت من قِبل عدد من الكتّاب كنماذج لمجتمعات واقعية محددة. وقد اقترح Mackenizie و Fisher (1922) الدالّة $\rho_u=\tanh(u^{-3/5})$ من أجل الارتباط بين المعدل الأسبوعي لهطول المطر في محطتين للأرصاد الجوية المسافة بينهما u ، كما اقترح Osborne (1942) الـدالّـة $\rho_u = e^{-\lambda u}$ في مسـوح إحصائية تتعلق بالغابات واستخدام

الأرض، واقترح Wold (1938) الدالّة $\rho_u=(l-u)/l$ في بعض أنواع السلاسل الزمنية الاقتصادية.

(٨ ـ ١٠) مجتمعات من الطبيعة

جرت تحرّيات تناولت مجتمعات متنوعة من الطبيعة. والبيان الإحصائي موصوف في الجدول (٨ـ٦). والدراسات الثلاث الأولى مأخوذة من خرائط. ويتألف المجتمع، في الدراسة الأولى، من 288 قراءة للارتفاع عن سطح البحر في مواقع متتالية تبعد عن بعضهما مسافة 0.1 ميل، وذلك في منطقة متموجة. والبيان الإحصائي في الدراستين التاليتين، هو الكسر من أطوال الخطوط المرسومة على خريطة

جدول (٨ ـ ٦) مجتمعات من الطبيعة استُخدمت في دراسات تتعلق بالمعاينة النمطية

نوع البيان الإحصائي	N	المرجع
الارتفاع عن سطح البحر مقروء في مواقع تبعد عن بعضهما 0.1 ميل على خريطة مساحة عسكرية.	288	Yates (1948), table 13
النسبة المئوية للمساحة (أ) أرض مزروعة، (ب) مكسوّة بالشجيرات، (ج) مكسوّة بالعشب، (د) مكسوّة بالغابات. وذلك على خطوط متوازية مرسومة على خريطة (Cover-type)	•	Osborne (1942)
النسبة المئوية للمساحة المغطاة بشجر التنّوب (Douglas Fir) على خطوط متوازية مرسومة على خريطة (Cover-type).	•	Osborne (1942)
درجة حرارة التربة (12 بوصة تحت العشب) على مدى 192 يومًا على التالي.	192	Yates (1948)
درجة حرارة التربة (4 بوصات تحت تربة عارية) على مدى 192 يومًا متتاليًا.	192	Yates (1948)
درجة حرارة الجوّ لفترة 192 يومًا.	192	Yates (1948)
إنتاج 96 صفًا من البطاطا.	96	Yates (1948)
حجم الخشب القابل للبيع في كل شريط، عرضه 3 chains، وبأطوال مختلفة (غابة Mt Stuart).	160	Finney (1948)
حجم خشب الـ Virgin لكل شريط، عرض 2.5 Chains، وطول 80 Chanis (غابة Black mountain).	288	Finney (1948)
حجم الخشب لكل شريط، عرض 2 Chains، وبأطوال مختلفة (غابة Dehra Dun).	292	Finney (1950)
عدد الشتول لكل مسكبة عرضها قدم واحدة في 4 مساكب لبذور الـ Hard Wood.	400†	Johnson (1943)
عدد الشتول لكل مسكبة عرضها قدم واحدة في 3 مساكب لبذور الصنوبر.	400†	Johnson (1943)
عدد الشتول لكل مسكبة عرضها قدم واحدة في 6 مساكب لشجيرات الصنوبر المزروعة.	400†	Johnson (1943)

• نظريًا، يكون N لا نهائيًا إذا أمكن تصوّر الخطوط اللامتناهية في دقتها.
† بصورة تقريبية، فالعدد يتغير من مسكبة إلى أخرى.

(Cover-type) الذي يقع في نوع غطاء معين (مثلاً العشب). ويمكن اعتبار هذين المثالين، الأكثر قربًا من التغير المستمر بالمعنى الرياضي للكلمة.

وقـد قامت الـدراسات الثلاث التالية على درجات الحرارة في 192 يومًا متتاليًا مقاسة: (ا) على عمق 12 بوصة تحت التربة، (ب) على عمق 4 بوصات تحت التربة، (ج) في الهواء. وتمثل هذه النسبة تدرّجًا في اتجاه التأثير المتزايد لتغيرات غريبة الأطوار في الطقس من يوم إلى آخر، وذلك بالمقارنة مع التأثيرات الفصلية البطيئة.

وتعالج الدراسات الباقية إنتاج شتول أو أغراس في نوع من التسلسل يقع على طول خط. وفي دراسـة تتعلق بالبـطاطا، وهي نموذجية في هذه المجموعة، يتألف المجتمع المنتهي من مجموع إنتاج 96 من الصفوف في حقل. وقد تتوافر بيانات إحصائية إضافية، باعتبار أن البحث عن مثل هذه البيانات لم يستنفد كل المراجع.

وفي بعض الدراسات يُقارن V_{sy} مع التباين V_{st2} في عيّنة عشوائية طبقية بطبقات حجمها $2k$ ووحدتين في كل طبقة. وهذه المقارنة مهمة لأنه يمكن الحصول على تقدير غير منحاز لِـ V_{st2} من بيان العيّنة. ولا يمكن القيام بذلك من أجل V_{st1} (أي طبقات حجمهـا k ووحـدة واحـدة من كل طبقة) أو من أجل V_{sy}. ويروي كتّاب آخرون مقارنات بين V_{sy} وكل من V_{st1} و V_{st2} ولا تقدم أغلب المصادر مقارنات مع V_{ran} في شكل جاهز للاستخدام، ولكن يبدو، بصورة عامة، أن V_{st2} قد أعطت كسبًا في الدقة فوق V_{ran}.

وفي أبحاث Yates و Finney أُعطيت مقارنات لمدى من القيم لِـ n و k ضمن كل مجتمـع منتـه. وفي هذه الحـالات نجد أن البيان الإحصائي في الجدول (٨-٧) هو المتوسطات الهندسية لنسب التباين الخاصة بقيم k كل بمفردها. ويقوم كتاب آخرون بالحسابات في حالة قيمة واحدة فقط لِـ k في كل مجتمع، ولكنهم قد يعطون بيانًا في حالة مفـردات مختلفـة أو في حالة عدة مجتمعات تتميز بالطبيعة نفسها وهنا أيضًا أُخذت المتوسطات الهندسية لنسب التباين.

ومع أن البيان الإحصائي محدود في مداه، فإن النتائج تدعو للإعجاب. وفي الدراسات التي تسمح بالمقارنة مع V_{st1} تُظهر المعاينة النمطية كسبًا دائمًا في الدقة يستحق

الاعتبار بالرغم من كونه متواضعًا. ووسيط النسب V_{st1}/V_{sy} هو 1.4 . والمكاسب كبيرة بالمقارنة مع V_{st2} حيث وسيط النسب 1.9 .

وتتفق الاتجاهات الضمنية للنتائج مع التوقعات، بالرغم من أنه يجب ألا يُعوّل عليها كثيرًا بالنظر إلى العدد الصغير من الدراسات. ويكون الكسب أكبر في أنواع من البيانـات الإحصـائية نستطيع أن نقول، تخمينًا، إن التغير فيها أقرب إلى أن يكون مستمرًا. والهبوط في V_{st1}/V_{sy} من درجة حرارة التربة إلى درجة حرارة الجو يمكن أن يكون متوقّعًا أيضًا من وجهة النظر هذه. وفي المفردات الثلاث الأخيرة (بيانات حول مشاتل لأغـراس شجـر الغابات)، فإن البيان الوحيد الذي لا يُظهر كسبًا هو ذلك المتعلق بمشاتل الصنوبر المزروعة زرعًا، وهي أقدم وأكثر انتظامًا من المشاتل المبذورة بذرًا.

جدول (٨-٧) الدقة النسبية للمعاينة النمطية والمعاينة العشوائية الطبقية

البيان الإحصائي	مدى k	الدقة النسبية للنمطية إلى الطبقية	
		V_{st2}/V_{sy}	V_{st1}/V_{sy}
الارتفاعات	2–20	5.68	2.99
النسبة المئوية للمساحة (4 cover-type)		4.42	—
النسبة المئوية للمساحة (Douglas Fir)		1.83	—
درجة حرارة التربة (21 بوصة)	2–24	4.23	2.42
درجة حرارة التربة (4 بوصات)	4–24	2.07	1.45
درجة حرارة الجوّ	4–24	1.65	1.26
البطاطا	3–16	1.90	1.37
حجم الخشب (Mt Stuart)	2–32	1.35	1.07
حجم الخشب (Black's Mt)	2–24	1.44	1.19
حجم الخشب (Dehra Dun)	2–32	1.89	1.39
مشاتل (Hard Wood)	14	1.89	—
مشاتل الصنوبر المبذورة	14–24	2.22	—
مشاتل الصنوبر المغروسة	12–22	0.93	—

(٨ - ١١) تقدير التباين من عيّنة بمفردها

من نتائج عيّنة عشوائية بسيطة فيها $n>1$ يمكن حساب تقدير غير منحاز لتباين متوسط عيّنة. وهذا التقدير غير منحاز بصرف النظر عن شكل المجتمع. وبما أنه يمكن اعتبار العيّنة النمطية كعيّنة عشوائية بسيطة فيها $n=1$، فإن هذه الخاصة المفيدة لا تصحّ في العيّنة النمطية. وكتوضيح، لنعتبر مثال «المنحني الجيبي». وليكن،

$$y_i = m + a \sin \frac{\pi i}{2}$$

حيث $k=4$ و $i=1,2,\ldots,4n$، فالملاحظات المتتالية في المجتمع هي:

$$(m+a), m, (m-a), m, (m+a), m, (m-a), m, \ldots$$

وإذا اختير $i=1$ كعنصر أول، فإن لجميع عناصر العيّنة النمطية القيمة $(m+n)$. وفي الاختيارات الثلاثة الممكنة لِـ i، يكون لجميع العناصر القيم $(m\text{-}a),m$ أو m، على الترتيب. وهكذا فإنه لا تتوافر لنا من عيّنة واحدة أية وسائل لإيجاد أو تقدير قيمة a. ولكن التباين الحقيقي لمتوسط عيّنة نمطية هو $a^2/2$. ويبيّن التوضيح استحالة القيام بتقدير غير منحاز للتباين في حال وجود تغير دوري.

وهذه النتائج لا تعني أنه لا يمكن القيام بشيء، فباستثناء حالة التغير الدوري، يمكن أن نعرف عن بنية المجتمع ما يكفي لتمكيننا من تطوير نموذج رياضي يمثل، بصورة مناسبة، نوع التغير الموجود. وقد نكون عندئذ قادرين على صياغة علاقة في هذا النموذج تعطي تقديرًا غير منحاز، على وجه التقريب، للتباين، وذلك بالرغم من أنه قد يكون منحازًا بصورة رديئة في نماذج أخرى. ويجب أن يستند قرار استخدام أحد هذه النماذج على حكمة الباحث.

ونوضح فيما يلي بعض النماذج البسيطة مع تقديرات التباين الموافقة. وسوف لا نعطي أية براهين.

وتنطبق النماذج الأبسط على مجتمعات تتألف فيها y_i من اتجاه مضافًا إلى مركبة عشوائية، مثل،

$$y_i = \mu_i + e_i$$

حيث μ_i دالة ما في i . ونفترض من أجل المركبة العشوائية وجود مجتمع فوقي يكون فيه :

$$\mathscr{E}(e_i) = 0, \qquad \mathscr{E}(e_i^2) = \sigma_i^2, \qquad \mathscr{E}(e_i e_j) = 0 \qquad (i \neq j)$$

وتعطي صيغة مقترحة لِـ $s_{sy}{}^2$ تقديرًا غير منحاز للتباين إذا كان ،

$$\mathscr{E}E(s_{sy}{}^2) = \mathscr{E}V_{sy}$$

أي إذا كان غير منحاز فوق جميع المجتمعات المنتهية التي يمكن سحبها من المجتمع الفوقي .

مجتمع بترتيب «عشوائي»

$$\mu_i = \text{ثابت} \qquad (i = 1, 2, \ldots, N)$$

$$s_{sy\,1}^2 = \frac{N-n}{Nn} \frac{\sum (y_i - \bar{y}_{sy})^2}{n-1} \tag{8.43}$$

وتنطبق هذه الحالة عندما نكون على ثقة من أن الترتيب، من أجل المفردات التي نقيسها، هو في الأساس عشوائي . وتبقى علاقة التباين هي العلاقة نفسها الموافقة لعيّنة عشوائية بسيطة، كما تكون غير منحازة إذا كان النموذج صحيحًا .

تأثيرات التقسيم إلى طبقات فقط

$$\mu_i \quad \text{ثابت} \qquad (rk + 1 \leq i \leq rk + k)$$

$$s_{sy\,2}^2 = \frac{N-n}{Nn} \frac{\sum (y_i - y_{i+k})^2}{2(n-1)} \tag{8.44}$$

وفي هذه الحالة، يبقى المتوسط ثابتًا ضمن كل طبقة فيها k من الوحدات . والتقدير s_{sy2}^2 ، المبني على متوسط مربعات الفروق المتتالية، هو تقدير منحاز. وهو يحوي مساهمة، لا نريدها، ناشئة عن الفرق بين المقادير μ في طبقتين متجاورتين، كما أن للطبقتين الأولى والأخيرة القليل من الوزن في تقدير المركبة العشوائية للتباين . وبفرض

أن النموذج صحيح ، ومع عيّنة كبيرة إلى حد ما ، قد يكون هذا التقدير، بصورة عامة ، تقديرًا مرتفعًا جدًّا .

اتجاه خطي

$$\mu_i = \mu + \beta i$$

$$s^2_{sy3} = \frac{N-n}{N}\frac{n'}{n^2}\frac{\sum (y_i - 2y_{i+k} + y_{i+2k})^2}{6(n-2)} \qquad (1 \le i \le n-2) \qquad (8.45)$$

والتقـــدير مبني على حدود تربيعية متتـــابعـــة في متـــوالية قيم y_i . ويحـوي مجمـوع المربعات $(n\text{-}2)$ من الحدود. وقد رأينا في حالة وجود اتجاه خطي (فقرة ٨ ـ ٧) أنه يمكن إلغـاء الاتجـاه باستخـدام تصحيحـات النهـاية. والحـد n'/n^2 هو مجمـوع مربعـات الترجيحات في $\bar{y}_{wsy}$. وما لم يكن n صغيرًا، فيمكن وضع العامل المعتاد $\frac{1}{n}$ بدلاً من n'/n^2 . ولأن الطبقتين المتطرفتين تتلقيان ترجيحات صغيرة جدًّا، فإن التقدير سيكون منحازًا ما لم يكن σ_i^2 ثابتًا، إلا أنه ينبغي أن يكون مُرضيًا إذا كان n كبيرًا وكان النموذج صحيحًا.

وفي حال وجود تغير مستمر لنوع أكثر تعقيدًا، فقد تعطي العلاقات السابقة نتائج غير دقيقة. وفي الجدول (٨ـ٨) طُبّقت العلاقتان الثانية والثالثة على ست مساكب لشتول شجر الغابات (Johnson, 1943) والعلاقة التربيعية أفضل بقليل من تلك المبنية على الفروق المتتالية، ولكن كلاً منهما تعطي مبالغات جدية في التقدير.

جدول (٨ـ٨) تباينات متوسط عدد الشتول في العيّنة (بيان *Johnson*)

	المسكبة	الفعلي V_{sy}	s^2_{sy2}	s^2_{sy3}
القيقب الفضي	1	0.91	2.8	2.5
	2	0.74	3.6	2.9
شجرة الدردار الأمريكي	1	4.8	28.4	12.6
	2	15.5	22.6	18.6
شجر أبيض من فصيلة الصنوبر	1	5.5	17.2	11.2
	2	2.0	11.6	6.4
أناناس أبيض	1	8.2	21.0	21.9

وقــد ناقش Osborne (1942) ، Cochran (1946) و Matern (1947) علاقـات استنبــطوهــا من فرضيات بسيطة حول طبيعــة مصــوّر الارتبــاط. وقـد تقصّى Yates (1949) تقديرًا مبنيًا على فروق d_u أولها:

$$d_1 = (\tfrac{1}{2}y_1' + y_3' + y_5' + y_7' + \tfrac{1}{2}y_9') - (y_2' + y_4' + y_6' + y_8') \qquad (8.46)$$

ويمكن أن يبـدأ الفرق الثاني، d_2 ، بِـ y'_9 ، وهلم جرًّا. وعندئذ نأخذ لتقدير تباين $\bar{y}_{sy}$ العبارة:

$$\hat{V}(\bar{y}_{sy}) = \frac{N-n}{Nn} \sum_{u=1}^{g} \frac{d_u^2}{7.5g} \qquad (8.47)$$

حيث العامل 7.5 هو مجموع مربعات المعاملات في أي d_u ، و g هو عدد الفروق التي تقـدمهـا العيّنـة (g يسـاوي تقـريبًا $n/9$). وفي المجتمعات من الطبيعة التي درسها Yates ، تتفوق صيغة كهذه على الصيغة s_{sy2}^2 المبنية على الفروق المتتالية، ولكنها لاتزال تبالغ في تقدير التباين الحقيقي لِـ $\bar{y}_{sy}$ وهو $V(\bar{y}_{sy})$.

والخلاصة، ليست هناك ندرة في الصيغ الخاصة بتقدير التباين، ولكنها تبدو جميعها وكأن مداها التطبيقي محدود.

وفي حالة $N=nk$ ، لنفرض أن n يقبل القسمة على عدد صحيح m (مثلاً 10). فالطريقة التالية تستخدم المعاينة النمطية، جزئيًا، وتقدّم تقدير عينة غير منحاز لِـ $V(\bar{y}_{sy})$. مبنيًا على (m-1) درجـة من الحـرية. لنسحب عيّنـة عشـوائية بسيطة حجمها m من وحدات المجتمع المرقمة من 1 إلى mk . ومع كل وحدة من هذه العيّنة، نأخذ أيضًا الوحدة الـ mk من كل mk من وحدات المجتمع التالية لها. وفي الواقع، ينقسم المجتمع بهذه الطريقة إلى mk عنقودًا حجم كل منها $N/mk=n/m$ ، ونختار عشوائيًا m عنقودًا، بحيث نحصل على عيّنة عشوائية بسيطة من m عنقودًا. وعلى سبيل المثال، لنفرض أننا نريد عيّنـة تشكـل 20% من مجتمـع فيه $n=2400$ ، بحيث يكون $n=480$ ، $k=5$. ولنـأخــذ عيّنـة عشــوائية بسيطة حجمهـا $m=10$ من وحـدات رُقّمت من 1 إلى $mk=50$ ، ولنـأخـذ مع كل وحـدة من هذه الـوحدات العشر الوحدة الخمسين من

كل 50 وحــدة تليهــا . ونحصــل عنـدئـذ على عيّنـة من عشرة عنـاقيد حجم كل منها 48=50 /2400 .

وقــد درس Gautschi (1957) دقـة هذه الـطريقـة تحت البنى المجتمعية التي افترضناها في هذا الفصل . وكما هو متوقع ، تقع الدقة بين دقة المعاينة العشوائية البسيطة ودقة المعاينة النمطية مع $m=1$.

(٨ ـ ١٢) المعاينة النمطية الطبقية

رأينا أنه إذا جرى ترتيب الوحدات بصورة مناسبة فإن المعاينة النمطية تقدم نوعًا من التقسيم إلى طبقات مع بقاء كسر المعاينة ثابتًا من طبقة إلى أخرى . وإذا قسمنا إلى طبقات وفقًا لقاعدة أخرى فيمكن سحب عينة نمطية منفصلة ضمن كل طبقة مع تحديد نقـطة البـدء ضمن كل طبقة بصورة مستقلة . وسيكون هذا مناسبًا إذا أردنا تقديرات منفصلة لكل طبقة أو إذا كنا سنستخدم كسور معاينة غير متساوية . وستكون هذه الطريقة أكثر دقة بالطبع من المعاينة العشوائية الطبقية إذا كانت المعاينة النمطية ضمن الطبقات أكثر دقة من المعاينة العشوائية البسيطة ضمن الطبقات .

وإذا كان $\bar{y}_{syh}$ متوسط العيّنة النمطية في الطبقة h ، فيكون تقدير متوسط المجتمع وتباين هذ التقدير:

$$\bar{y}_{stsy}=\sum W_h\bar{y}_{syh}, \qquad V(\bar{y}_{stsy})=\sum W_h^2 V(\bar{y}_{syh})$$

ومع عدد قليل من الطبقات فقط تؤدي مسألة إيجاد تقدير عينة لهذه الكمية إلى مسألة نوقشت سابقًا وهي مسألة إيجاد تقدير عيّنة مُرضٍ لِـ $V(\bar{y}_{syh})$ في كل طبقة .

وعندما يكون عدد الطبقات كبيرًا، فقد يكون استخدام تقدير مبنيّ على طريقة «الطبقات المنهارة» (فقرة ١٥ـ١٢) مفضلاً . ومن نتائج تلك الفقرة نستنتج أن التقدير

$$v(\bar{y}_{stsy})=\sum' W_h'^2(\bar{y}_{syh}-\bar{y}_{syj})^2 \qquad (8.48)$$

حيث يمتد المجموع فوق أزواج الطبقات، هو في المتوسط تقدير بالزيادة، حتى في حال وجـود تغيّر دوري ضمن الطبقات . ويمكن الحصول على تقدير غير منحاز لتباين الخطأ إذا سحبنا ضمن كل طبقة عينتين نمطيتين مع بدايتين عشوائيتين مختلفتين

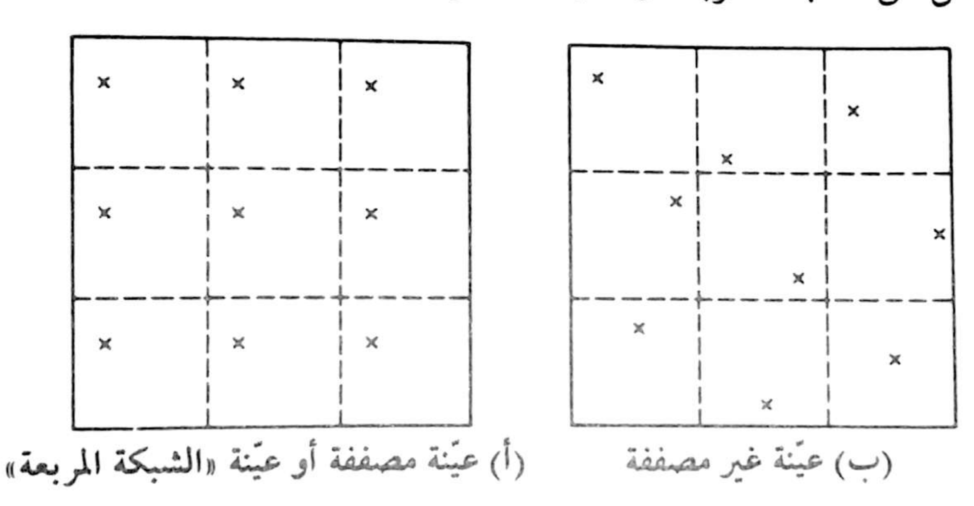
وفـترة طولهـا $2k$ ، وتمدّنا كل طبقة بدرجة واحدة من الحرية . وسيكون هناك بعض الخسارة في الدقة إذا كانت المعاينة النمطية فعّالة . وإذا كان هناك العديد من الطبقات فيمكن استخـدام عيّنـة نمطية واحدة من كل منها أو من معظمها، وفي هذه الحالة نسحب عيّنة جزئية عشوائية من وحدتين في كل طبقة لأغراض تقدير الخطأ .

(٨ ـ ١٣) المعاينة النمطية ذات البعدين

عند معاينة مساحة ، نجد أن أبسط تمديد للعينة النمطية ذات البعد الواحد هو نموذج الشبكة المربعة المبيّن في الشكل (٨ ـ ٤ ا) . وتتحدد العيّنة تمامًا باختيار زوج من الأعداد العشوائية لتثبيت إحداثيات الوحدة العليا على اليسار . وقد دُرس أداء الشبكة المربعة في مجتمعات نظرية ومجتمعات من الطبيعة . وتقصّى Matern (1960) أفضل نوع من المعاينة عندما يكون الارتباط بين أي نقطتين من المساحة دالة في المسافة الفاصلة بينهما d ، متناقصة باطراد ومقعرة إلى الأعلى . وفي حالة مصوّرات ارتباط مثل $e^{-\lambda d}$ تقوم الشبكة المربعة بعمل جيد، وتتفوق على المعاينة العشوائية البسيطة أو الطبقة بوحدة واحدة من كل طبقة، هذا بالرغم من أن Matern يعلّل توقعه بأن أفضل نموذج لهذه الحالة هو الشبكة المثلثة التي تقع النقاط فيها على رؤوس مثلثات متساوية الأضلاع .

وفي أربع عشرة تجربة زراعية خاضعة للشروط نفسها، وجد Haynes (1948) أن دقة الشبكة المربعة مساوية تقريبًا لدقة معاينة عشوائية بسيطة ذات بعدين . وقد درس Milne (1959) الشبكة المربعة المركزية، التي تقع النقطة فيها في مركز المربع ، وذلك في 50 من التجارب الخاضعة للشروط نفسها. وكان أداؤها أفضل من المعاينة العشوائية البسيطة، وربما أفضل بقليل من المعاينة العشوائية الطبقية، مع أن هذا الفرق لم يكن مهمًا من الناحية الإحصائية . وتقترح هذه النتائج أن تأثيرات الارتباط الذاتي ضعيفة في هذا النوع من البيانات الاحصائية، على الأقل . ولتقدير المساحة المغطاة بالغابات أو بالماء على خريطة ، وجد Matern أن الشبكة المربعة متفوقة على الطريقة العشوائية في مثالين .

ويبينّ الشكل (٨ ـ ٤ ب) عيّنة نمطية بديلة تدعى العيّنة غير المصفّفة. ونختار هنا إحداثيات الوحدة اليسرى العليا أولاً بوساطة زوج من الأعداد العشوائية. ويحدّد عددان عشوائيان إضافيان الإحداثيين الأفقيين للوحدتين الباقيتين في العمود الأول من الطبقات. ونحتاج إلى عددين عشوائيين آخرين لتثبيت الإحداثيات الشاقولية للوحدتين الباقيتين في الصف الأول من الطبقات. وعندئذ تحدد الفترة الثابتة k (وهي تساوي أضلاع المربعات) مواقع جميع النقاط. وتشير تحرّيات Quenouille (1949) و Das (1950) لمصوّري ارتباط بسيطين وثنائيي البعد أن النموذج غير المصفّف سيتفوق غالبًا على كل من الشبكة المربعة والمعاينة العشوائية الطبقية.



(أ) عيّنة مصففة أو عيّنة «الشبكة المربعة» (ب) عيّنة غير مصففة

شكل (٨ـ٤) نوعان من العيّنات النمطية ذات البعدين

ونحصل على دليل إضافي يشير إلى تفوق عينة غير مصففة من الخبرة في تصميم التجارب، حيث وُجد المربع اللاتيني كطريقة دقيقة لترتيب معالجات في حقل مستطيل. ويمكن اعتبار المربع اللاتيني 5×5 في الشكل (٨ ـ ٥ ا) كتقسيم للحقل إلى خمس عيّنات نمطية، واحدة من أجل كل حرف. وهناك بعض الدلالة على أن ذلك المربع الخاص، الذي يُدعى حركة الفارس أكثر دقة بقليل من مربع 5×5 نختاره عشوائيًا. وربما كان ذلك بسبب غياب التصفيف في الأقطار وفي الصفوف والأعمدة على حدّ سواء.

وقد استخدم Homeyer و Black (1946) مبدأ المربع اللاتيني لمعاينة حقول من الشوفان مستطيلة الشكل، وكل حقل يحوي 21 وحدة تجريبية. وفي الشكل

(٨ ـ ٥ ب) نرمز للعيّنات النمطية الثلاث بالأحرف A ، B ، و C على التوالي . ومع اختيار أحد هذه الأحرف عشوائيًا في كل حقل، أعطى هذا الترتيب الناتج زيادة 52% تقريبًا، في الدقة فوق المعاينة العشوائية الطبقية التي تعتبر الصفوف كطبقات . ولا يحقق الترتيب خاصة المربع اللاتيني تمامًا، لأن كل حرف يظهر 3 مرات في أحد الأعمدة ومرتين في الأعمدة الأخرى، ولكنه يقترب من هذه الخاصة بالقدر الممكن .

A	B	C	D	E
D	E	A	B	C
B	C	D	E	A
E	A	B	C	D
C	D	E	A	B

(ا) المربع اللاتيني «حركة الفارس»

A	B	C
B	C	A
C	A	B
A	B	C
B	C	A
C	A	B
A	B	C

(ب) تصميم نمطي لحقل مستطيل 3×7

شكل (٨ ـ ٥) تصميمان نمطيّان مبنيّان على المربع اللاتيني

ويناقش Yates (1960) ، الذي يصطلح على تسمية هذا النوع من الترتيبات بالمعاينة الشبكية ، استخدامها في معاينة ذات بعدين أو ثلاثة أبعاد . وفي حالة ثلاثة أبعاد يمكن تمثيل كل صف وعمود ومسافة شاقولية في العيّنة باختيار p من الوحدات من بين الـ p^3 وحدة في المجتمع . ومع p^2 وحدة في العيّنة يمكن تمثيل كل من تراكيب المستويات الـ p^2 للصفوف والأعمدة، وللصفوف والمسافات الشاقولية، وللأعمدة والمسافات الشاقولية . وقد تقصّى Patterson (1954) الترتيبات التي تعطي تقديرًا غير منحاز للخطأ .

(٨ ـ ١٤) خلاصة

من السّهل سحب العيّنات النمطية وتنفيذها . وفي معظم الدراسات التي ذكرناها في هذا الفصل، والتي تناولت كلًّا من المجتمعات الاصطناعية والمجتمعات من الطبيعة، كانت المقارنة في الدقة مع العيّنات العشوائية الطبقية لصالح العيّنات النمطية . وكانت مساوئها هي أنها يمكن أن تعطي دقة رديئة عند وجود دورية لا مجال للشك فيها، وأننا لا نعرف طريقة موثوقة لتقدير $V(\bar{y}_{sy})$ من بيان العيّنة .

(٨ ـ ٥ ب) نرمز للعيّنات النمطية الثلاث بالأحرف A ، B ، و C على التوالي. ومع اختيار أحد هذه الأحرف عشوائيًا في كل حقل، أعطى هذا الترتيب الناتج زيادة 52% تقريبًا، في الدقة فوق المعاينة العشوائية الطبقية التي تعتبر الصفوف كطبقات. ولا يحقق الترتيب خاصة المربع اللاتيني تمامًا، لأن كل حرف يظهر 3 مرات في أحد الأعمدة ومرتين في الأعمدة الأخرى، ولكنه يقترب من هذه الخاصة بالقدر الممكن.

A	*B*	*C*	*D*	*E*
D	*E*	*A*	*B*	*C*
B	*C*	*D*	*E*	*A*
E	*A*	*B*	*C*	*D*
C	*D*	*E*	*A*	*B*

(ا) المربع اللاتيني «حركة الفارس»

A	*B*	*C*
B	*C*	*A*
C	*A*	*B*
A	*B*	*C*
B	*C*	*A*
C	*A*	*B*
A	*B*	*C*

(ب) تصميم نمطي لحقل مستطيل 3×7

شكل (٨ ـ ٥) تصميهان نمطيّان مبنيّان على المربع اللاتيني

ويناقش Yates (1960)، الذي يصطلح على تسمية هذا النوع من الترتيبات بالمعاينة الشبكية، استخدامها في معاينة ذات بعدين أو ثلاثة أبعاد. وفي حالة ثلاثة أبعاد يمكن تمثيل كل صف وعمود ومسافة شاقولية في العيّنة باختيار p من الوحدات من بين الـ p^3 وحدة في المجتمع. ومع p^2 وحدة في العيّنة يمكن تمثيل كل من تراكيب المستويات الـ p^2 للصفوف والأعمدة، وللصفوف والمسافات الشاقولية، وللأعمدة والمسافات الشاقولية. وقد تقصّى Patterson (1954) الترتيبات التي تعطي تقديرًا غير منحاز للخطأ.

(٨ ـ ١٤) خلاصة

من السّهل سحب العيّنات النمطية وتنفيذها. وفي معظم الدراسات التي ذكرناها في هذا الفصل، والتي تناولت كلًّا من المجتمعات الاصطناعية والمجتمعات من الطبيعة، كانت المقارنة في الدقة مع العيّنات العشوائية الطبقية لصالح العيّنات النمطية. وكانت مساوئها هي أنها يمكن أن تعطي دقة رديئة عند وجود دورية لا مجال للشك فيها، وأننا لا نعرف طريقة موثوقة لتقدير $V(\bar{y}_{sy})$ من بيان العيّنة.

وفي ضوء هذه النتائج يبدو أنه يمكننا، باطمئنان، أن نوصي باستخدام المعاينة النمطية في الحالات التالية:

١ ـ عنـدمـا يكون ترتيب المجتمع، في الأساس، عشوائيًا، أو أنه يحوي، في أسوأ الحالات، صورة معتدلة من الانقسام إلى طبقات. وتُستخدم المعاينة النمطية هنا لسهولتها، مع القليل من توقع الكسب في الدقة. وتتوافر لنا تقديرات عيّنة للخطأ تتصف، إلى حد معقول، بأنها غير منحازة (الفقرة ٨ ـ ١١).

٢ ـ حيث يُستخدم التقسيم إلى عدد كبير من الطبقات، وتُسحب عيّنة نمطية مستقلة من كل طبقة. وفي هذه الحالة تميل تأثيرات أي من الاتجاهات الدورية المستترة إلى أن تلغي بعضها بعضًا. ويمكن الحصول على تقدير للخطأ يتصف بالمبالغة (أي تقدير بالزيادة)، (فقرة ٨ ـ ١٢). ويمكن، بصورة بديلة، استخدام نصف عدد الطبقات وسحب عيّنتين نمطيتين، مع بدايات عشوائية مستقلة، من كل طبقة، وتعطي هذه الطريقة تقديرًا غير منحاز للخطأ.

٣ ـ في معـاينـة جزئية لوحـدات عنقودية (الفصل العاشر). وفي معظم التطبيقات العملية الموافقة لهذه الحالة يمكن الحصول على تقدير لخطأ المعاينة غير منحاز أو غير منحاز تقريبًا. واستخدام المعاينة النمطية هو أمر شائع في هذه الحالة.

٤ ـ عند معاينة مجتمعات تتصف بتغير من النوع المستمر، شريطة ألا يكون الحصول على تقـدير لخطأ المعاينة مطلوبًا بصورة منتظمة. وإذا قمنا بسلسلة من المسوح الإحصائية من هذا النوع، فقد يكون تدقيق أخطاء المعاينة هذه من وقت لآخر كافيًا. وقد بيّن Yates (1948) كيف يمكن القيام بذلك، وذلك بأخذ ملاحظات إضافية.

تماريـن

(٨ ـ ١) البيان الإحصائي أدناه هو عدد الشتول في كل قدم من مسكبة طولها 200 قدم. أحسب تباين متوسط عيّنة نمطية مؤلفة من القدم العشرين من كل عشرين قدمًا. وقارنه مع التباين في حالة ا) عيّنة عشوائية بسيطة، ب) عيّنة عشوائية طبقية بوحدتين من كل طبقة، ج) عيّنة عشوائية طبقية بوحدة واحدة من كل طبقة. في جميع العيّنات نأخذ $n = 10$. [$\sum (y_i - \bar{Y})^2 = 23{,}601$.]

عدد الشتول

	1–20 1	21–40 2	41–60 3	61–80 4	81–100 5	قدم 101–120 6	121–140 7	141–160 8	161–180 9	181–200 10	مجاميع العيّنة النمطية
	8	20	26	34	31	24	18	16	36	10	223
	6	19	26	21	23	19	13	12	8	35	182
	6	25	10	27	41	28	7	8	29	7	188
	23	11	41	25	18	18	9	10	33	9	197
	25	31	30	32	15	29	11	12	14	12	211
	16	26	55	43	21	24	20	20	13	7	245
	28	29	34	33	8	33	16	17	18	6	222
	21	19	56	45	22	37	9	12	20	14	255
	22	17	39	23	11	32	14	7	13	12	190
	18	28	41	27	3	26	15	17	24	15	214
	26	16	27	37	4	36	20	21	29	18	234
	28	9	20	14	5	20	21	26	18	4	165
	11	22	25	14	11	43	15	16	16	4	177
	16	26	39	24	9	27	14	18	20	9	202
	7	17	24	18	25	20	13	11	6	8	149
	22	39	25	17	16	21	9	19	15	8	191
	44	21	18	14	13	18	25	27	4	9	193
	26	14	44	38	22	19	17	29	8	10	227
	31	40	55	36	18	24	7	31	8	5	255
	26	30	39	29	9	30	30	29	10	3	235
مجاميع الطبقات	410	459	674	551	325	528	303	358	342	205	4155

(٨ ـ ٢) رُتّب مجتمـع من 360 أسرة (مرقمة من 1 إلى 360) في بلتيمور ترتيبًا أبجديًا وفقًا لكنية معيل الأسرة ووضع في ملفّ. ووقعت الأسر التي لم يكن معيلوها من البيض عنـد الأرقـام التـالية: 28 ,31-33 ,36-41 ,44 ,45 ,47 ,55 ,56 ,58 ,68 ,69 ,82, 83 ,85 ,86 ,89-94 ,98 ,101-99 ,107-110 ,114 ,154 ,156 ,178 ,223 ,224 ,296 ,298-300, 302-304 ,306-323 ,325-331 ,333 ,335-339 ,341 ,342 . (وتُظهر الأسر غير البيضاء بعضًا من «التكتّل» بسبب وجود علاقة بين الكنية واللون).

قارن دقة 1 من 8 عيّنة نمطية، مع عيّنة عشوائية بسيطة من الحجم نفسه، وذلك لتقدير نسبة الأسر التي لا يكون معيلها أبيض.

(٨ ـ ٣) يتـألف جوار من ثلاث جاليات مكتـظة مؤلفـة من أشخـاص يعود أسـلافهم، على الـترتيب، إلى أصل أنجلو ـ ساكسوني أو بولوني أو إيطالي. ويوجد فهرس حديث يرد فيه الأشخاص ضمن منزل وفقًا للترتيب التالي: الزوج، الزوجة، الأطفـال (حسب أعـمارهم)، آخـرون. وقـد رُتّبت المنـازل وفقـا لموقعها على طول الشوارع. ومتوسط عدد أفراد الأسرة هو خمسة.

والاختيار هو بين عيّنة نمطية تأخذ الشخص الخامس من كل خمسة أشخاص في الفهرس وبين 20% عيّنة عشوائية بسيطة. من أجل أي من المتغيرات التالية تتوقع أن تكون العيّنة النمطية أكثر دقة؟ ا) نسبة الأشخاص من أصل بولوني، ب) نسبة الذكور، ج) نسبة الأطفال. أعط أسبابًا لأجوبتك.

(٨ ـ ٤) في فهرس لثلاثة عشر منزلاً في شارع كانت قوائم الأشخاص كما يلي: M = ذكر بالغ، F = أنثى بالغة، m = طفل، f = طفلة

المنزل

1	2	3	4	5	6	7	8	9	10	11	12	13
M	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*	*M*
F	*F*	*F*	*F*	*F*	*F*	*F*	*F*	*F*	*F*	*F*	*F*	*F*
f	*f*	*m*		*m*	*f*	*f*	*m*	*m*	*m*	*f*	*f*	
m	*m*	*f*		*m*	*m*	*f*	*f*		*f*	*m*		
f	*f*			*f*		*m*						

قارن التباينات التي تعطيها 1 من 5 عيّنة نمطية و 20% عيّنة عشوائية بسيطة لتقدير:

أ) نسبة الذكور، ب) نسبة الأطفال وج) نسبة الأشخاص الذين يعيشون في أُسر مهنية (الأسر 1 ,2 ,3 ,12 و 13 تتصف بأنها أُسر مهنية). هل تدعم النتـائج هنا أجوبتك للتمرين (٨ - ٣)؟ في العيّنة النمطية، رقّم كل عمود من الأعلى إلى الأسفل ثم انتقل إلى أول العمود الذي يليه.

(٨ - ٥) في التمرين (٨ - ١) يمكن تقدير $V(\bar{y}_{sy})$ من خلال ا) اعتبار كل عيّنة نمـطية كعيّنة عشوائية بسيطة، ب) الادعاء بأن كل 1 من 20 عيّنة نمطية مؤلفة من عيّنتين على أساس 1 من 40 مع نقطتي بدء عشوائيتين منفصلتين. في كل طريقة، قارن متوسط التباينات المقدّرة مع التباين الفعلي لِـ $\bar{y}_{sy}$.

(٨ - ٦) في مجتمع ينطوي على اتجاه خطي (فقرة ٨ - ٦) بيّن أن العيّنة النمطية أقل دقة من العيّنة العشوائية الطبقية، بطبقات حجمها $2k$ ووحدتين من كل طبقة، إذا كان $n>[(4k+2)/(k+1)]$.

(٨ - ٧) يمكن تمثيل مجتمع ذي بعدين مع اتجاه خطي بالعلاقة:

$$y_{ij} = i + j \qquad (i, j = 1, 2, ,\ldots, nk)$$

حيث y_{ij} القيمة في الصف i والعمود j . ويتضمن المجتمع $N^2=n^2k^2$ من الوحدات.

اخترنا عيّنة نمطية على شكل شبكة مربعة بأن سحبنا عشوائيًا إحداثيي بداية مستقلين i_0, j_0 ، كل منهما بين 1 و k . وتتضمن العيّنة التي حجمها n_2 جميع الوحدات التي تكون إحداثياتها من الشكل:

$$i_0 + \gamma k, j_0 + \delta k$$

حيث γ ، δ أي عددين صحيحين بين 0 و $(n\text{-}1)$ بما فيه الطرفان.

بيّن أن لمتوسط هذه العيّنة نفس دقة متوسط عيّنة عشوائية بسيطة حجمها n^2 .

(٨ - ٨) إذا جرت المقارنة في التمرين (٨ - ٧) من أجل مجتمع ذي ثلاثة أبعاد مع اتجاه خطي فما هي النتيجة التي تتوقعها.

(٨ - ٩) من أجـل مجتمع فيه $y_i=i^2$ $(i=1,2,\ldots,16)$ قارن قيم $E(\hat{\bar{Y}} - \bar{Y})^2$ التي تعـطيهـا 1 من كل k عيّنـة نمـطية وطـرق Yates ، Sethi وSingh وآخـرين، وذلـك في حالـة $N=16, k=4, n=4$.

الفصل التاسع

المعاينة العنقودية
وحيدة المرحلة: عناقيد متساوية الحجم

(٩-١) دواعي المعاينة العنقودية

أشرنا عدة مرات في الفصول السابقة إلى مسوح إحصائية تتألف وحدة المعاينة فيها من زمرة أو عنقود من الوحدات الأصغر التي سميناها «عناصر» أو «وحدات جزئية». ويوجد سببان رئيسان للتطبيق الواسع الانتشار للمعاينة العنقودية. ومع أننا قد نتجه في البداية إلى استخدام العناصر كوحدات عيّنة، إلا أنه، في كثير من المسوح الإحصائية، لوحظ عدم توافر قائمة موثوقة بعناصر المجتمع، وإن وضع قائمة كهذه قد يكون مكلفًا إلى درجة تجعلها بعيدة المنال. وفي العديد من البلدان لا توجد قوائم كاملة وحديثة للسكان، أو للمنازل، أو للمزارع في أي منطقة جغرافية كبيرة. وباستخدام خرائط للمنطقة يمكننا، على أي حال، تقسيمها إلى وحدات مساحية مثل جادّات في المدن، أو قطاعات من الأرض حدودها قابلة للتعريف بسهولة في الأجزاء الريفية. وغالبًا ما يجري اختيار هذه العناقيد في الولايات المتحدة لأنها تحل مشكلة وضع قائمة بوحدات المعاينة.

وحتى عندما تتوافر قائمة بالمنازل كل بمفرده، فقد تشير اعتبارات اقتصادية إلى اختيار وحدة عنقودية أكبر. ومن أجل حجم معطى للعيّنة، تعطى الوحدة الصغيرة عادة نتائج أكثر دقة من الوحدة الكبيرة. وعلى سبيل المثال، تغطي عينة عشوائية بسيطة من 600 منزل بلدة ما تغطية أقل انحيازًا من 20 جادة، متوسط عدد المنازل في كل منها 30 منزلًا. إلا أن تحديد مواقع 600 منزل والانتقال فيما بينها سيتطلب من التكاليف الميدانية أكثر مما يتطلبه تحديد مواقع 20 جادة وزيارة جميع المنازل في كل جادة منها. وعندما نوازن بين التكلفة والدقة، فقد تثبت الوحدة الأكبر تفوقها.

ويمكن القيام باختيار عقلاني بين نوعين أو حجمين للوحدة من خلال المبدأ المألوف في اختيار الوحدة التي تعطي التباين الأصغر من أجل تكلفة محددة، أو التكلفة الأقل من أجل تباين محدد. وكما في العديد من القرارات التطبيقية، فقد تكون هناك عوامل غير منظورة إذ قد يكون لنوع ما من الوحدات بعض الميزات أو المساوىء التي يصعب تناولها في حساب التكاليف. وعند معاينة محصول في طور النمو، تقترح بعض الخبرات أن الوحدة الصغيرة يمكن أن تعطي تقديرات منحازة بسبب الريبة في دقة تعيين حدود الوحدة. وقد وجد Homeyer و Black (1946) أن وحدات قياسها 3×3 أقدام أعطت إنتاجًا من الشوفان أعلى بحوالي 8% من وحدات قياسها 2×2 قدمًا، وربما كان ذلك بسبب أن العاملين في المعاينة يميلون إلى اعتبار السنابل الواقعة على الحدود والمشكوك في أمرها سنابل ضمن الوحدة. ويذكر Sukhatme (1947) نتائج مشابهة عند معاينة حقول من القمح أو الأرز.

ويعالج هذا الفصل الحالة التي تتضمن فيها كل وحدة عنقودية العدد نفسه من العناصر أو الوحدات الجزئية.

(٩ ـ ٢) قاعدة بسيطة

عندما تكون المسألة هي مقارنة عدد قليل من حجوم محددة للوحدات أو من أنواع محددة منها، فإن النتيجة التالية تكون مفيدة.

نظرية (٩-١)

ليكن

الحجم النسبي للوحدة = M_u

التباين بين مجاميع الوحدات = S_u^2

التكلفة النسبية لقياس وحدة واحدة = C_u

فعندئذ تكون التكلفة النسبية من أجل تباين محدد، أو التباين النسبي من أجل تكلفة محددة متناسبًا مع $C_u S_u^2/M_u^2$ ، وينطبق هذا على معاينة عشوائية بسيطة يمكن فيها إهمال الـ(ت م م).

برهـان

لنفـرض أن V هو التبـاين المحـدّد لتقدير مجموع المجتمع . ففي النوع u من الوحدات يكون التقدير $N_u\bar{y}_u$ وتباينه :

$$V=\frac{N_u^2S_u^2}{n_u}; \qquad n_u=\frac{N_u^2S_u^2}{V} \qquad (9.1)$$

وتكلفة أخذ n_u من الوحدات هي $C_un_u = C_uN_u^2S_u^2/V$ وبما أن N_uM_u يبقى ثابتًا في أنواع مختلفة من الوحدات فتكون التكلفة متناسبة مع $C_uS_u^2/M_u^2$. وعلى الوجه الآخر، إذا كانت التكلفة C محددة فإن $n_u=C/C_u$ ومن (9.1) نجـد أن V متـنـاسـب مع $C_uS_u^2/M_u^2$ . . وهو المطلوب .

نتيجة ١

إذا عرّفنا الدقة النسبية الصافية لوحدة بأنها تتناسب عكسيًا مع التباين الذي نحصل عليه من أجل تكلفة مثبتة، فيمكن عرض النظرية (٩ ـ ١) على الشكل،

$$\text{الدقة النسبية الصافية متناسبة مع } \frac{M_u^2}{C_uS_u^2} \qquad (9.2)$$

نتيجة ٢

في تحليل التبـاين، غالبًـا ما تُحسب التبـاينات لوحدات مختلفة الحجوم، على أسـاس يُدعى الأسـاس المشترك ـ وهو في العادة ذلك الأساس القابل للتطبيق على الـوحـدة الأصغر. ولوضع التباينات على أساس مشترك نقسم التباين S_u^2 بين مجاميع وحدات حجمها M_u على M_u . ليكن،

$$\text{التباين بين مجاميع الوحدات (على أساس مشترك)} = S_u'^2=\frac{S_u^2}{M_u}$$

$$\text{التكلفة النسبية لأخذ حجم معطى للعيّنة} = C_u'\propto\frac{C_u}{M_u}$$

فعندئذ يمكن عرض النظرية (٩ ـ ١) كما يلي

$$\text{التكلفة النسبية من أجل الدقة نفسها} \propto\frac{C_uS_u^2}{M_u^2}\propto C_u'S_u'^2 \qquad (9.3)$$

$$\text{الدقة النسبية الصافية} \propto \frac{1}{C_u' S_u'^2} \quad (9.4)$$

وإذا تجاهلنا الفروق في تكاليف أخذ عيّنة (أي إذا كان C_u ثابتًا)، تكون الدقة النسبية الصافية للوحدة u متناسبة مع $\frac{1}{S_u'^2}$. ولذلك تكون عوامل أثر التصميم (deff) لِـ Kish في الوحدات المختلفة (فقرة ٤-١١) متناسبة مع $S_u'^2 = s_u^2/M_u$.

مثال

يقدم بيان Johnson (1941)، المتعلق بمسكبة من أغراس الصنوبر الأبيض، مثالاً بسيطًا. فقد احتوت المسكبة ستة صفوف طول كل منها 434 قدمًا. وهناك العديد من الطرق التي يمكن تقسيم المسكبة بموجبها إلى وحدات معاينة. ويبيّن الجدول (٩-١) بيانًا من أجل أربعة أنواع من الوحدات. وبما أن المسكبة أحصيت بالكامل فإن البيان يعطي قيمًا صحيحة للمجتمع.

وكانت الوحدات:

قدمًا واحدة من صف بمفرده.

قدمين من صف بمفرده.

قدمًا واحدة من عرض المسكبة.

قدمين من عرض المسكبة.

جدول (٩-١) بيان إحصائي لأربعة أنواع من وحدات المعاينة

بيان إحصائي تمهيدي	نوع الوحدة			
	قدم واحدة صف	٢ قدم صف	قدم واحدة مسكبة	٢ قدم مسكبة
الحجم النسبي للوحدة = M_u	1	2	6	12
عدد الوحدات في المجتمع = N_u	2604	1302	434	217
تباين المجتمع لكل وحدة = S_u^2	2.537	6.746	23.094	68.558
عدد الأقدام المتتالية التي يمكن إحصاؤها في 15 دقيقة	44	62	78	108

وقد افتُرض في الوحدتين الأولى والثانية أن المعاينة يمكن أن تكون طبقية وفقًا للصفوف، بحيث يمثل الـ S_u^2 تباينات ضمن الصفوف. كما افتُرضت المعاينة العشوائية البسيطة في الوحدتين الأخيرتين.

وبما أن التكلفة الرئيسة تقع في تعيين مواضع الوحدات ثم تعدادها فقد قُدّرت التكاليف وفقًا لدراسة زمنية (الصف الأخير من الجدول ٩ ـ ١). وفي حالة العيّنات الأكبر، يمكن تعداد قدر كبير من العيّنة في 15 دقيقة، لأن الوقت اللازم للتحرك من وحدة إلى أخرى يصبح أقل.

والكمية التي نريد تقديرها هي مجموع الأغراس الكلي في المسكبة. ووفقًا لرموز النظرية (٩ ـ ١)، يعطي الجدول (٩ ـ ١) قيم M_u و S_u^2. والقيم النسبية لـ C_u معبراً عنها بدلالة الزمن اللازم لتعداد وحدة واحدة، هي كما يلي:

	قدم واحدة مسكبة	٢ قدم مسكبة	قدم واحدة صف	٢ قدم صف
C_u (في فترة 15 دقيقة)	$\frac{12}{108}$	$\frac{6}{78}$	$\frac{2}{62}$	$\frac{1}{44}$

وبالاستناد إلى النظرية (٩ـ١)، نتيجة (١)، تمّ حساب قيم الدقة النسبية الصافية في الجدول (٩ـ٢).

جدول (٩ ـ ٢) الدقة النسبية الصافية للوحدات الأربع

	قدم واحدة/مسكبة	٢ قدم/مسكبة	قدم واحدة/صف	٢ قدم/صف
$\frac{M_u^2}{C_u S_u^2}$	$\frac{(144)(108)}{(12)(68.558)} = 18.90$	$\frac{(36)(78)}{(6)(23.094)} = 20.27$	$\frac{(4)(62)}{(2)(6.746)} = 18.38$	$\frac{44}{2.537} = 17.34$
	109	117	106	100

والتباينات بين الوحدات، معبراً عنها بدلالة أساس مشترك جديرة بأن يُنظر إليها أيضًا. فالقيم $S_u'^2 = S_u^2/M_u$، مطبقة على قدم واحدة من صف، هي، على الترتيب، 2.537, 3.373, 3.849 و 5.713 . ونلاحظ أن هذه التباينات تتزايد باطراد مع تزايد حجم الوحدة. وهذه النتيجة هي النتيجة الشائعة (مع أنه قد تقع استثناءات).

وبما أن الدقة النسبية الصافية متناسبة مع $1/C_u'S_u'^2$ ، فإن تكلفة أخذ حجم معطى للعينة يجب أن يتناقص في الوحدات الكبيرة، إذا كان لهذه الوحدات أن تُثبت اقتصاديتها .

وتبقى النظرية (٩ ـ ١) ونتيجتاها صحيحة في معاينة طبقية بمحاصّة تناسبية، وذلك إذا كانت جميع الطبقات من الحجم نفسه وكان S_u^2 و $S_u'^2$ يمثلان متوسطي التباينات ضمن الطبقات، ذلك، لأنه تحت الشروط المعروضة، يكون تباين تقدير مجمـوع المجتمع، متجاهلين عامل ال ت م م، مساويًا لِـ $N^2S_u^2/n$ ، وبالتالي فهو يتّخذ الشكل نفسه، كما في حالة معاينة عشوائية بسيطة . ولا تصح النظرية (٩ ـ ١) في أنواع من المعاينة أكثر تعقيدًا .

والمقصود من النتائج السابقة هو مجرّد توضيح للطريقة العامّة . ينبغي *دائمًا القيام بالمقارنات بين الوحدات من أجل النوع من المعاينة الذي سنستخدمه عمليًا، أو إذا لم يكن قد اتُّخذ قرار بذلك، فمن أجل الأنواع الداخلة في اعتبارنا* . والتغيرات في طريقة المعاينة أو التقدير، ستغيّر الدقة النسبية الصافية للوحدات المختلفة . وحتى مع طريقة مثبتة في التقدير والمعاينة، ستتغير الدقة النسبية الصافية مع حجم العيّنة إذا لم تكن التكلفة دالة خطية في الحجم، أوْ إذا كان الحجم من الكبر بحيث يجب أخذ عامل الـ ت م م بعين الاعتبار.

وتتنـاول الـدراسة عادة أكثر من مفردة واحدة . وإحدى الطرق هي أن نثبّت التكلفة الكلية، ونحسب الدقة النسبية لكل نوع من الوحدات ولكل مفردة . وما لم يكن أحـد أنـواع الـوحدات متفوقًا بصورة منتظمة، فإننا نتخذ قرار تسوية، يعطي ترجيحة رئيسة للمفردات الأكثر أهمية .

ونظرًا للعدد الكبير من العوامل التي تؤثر في النتائج، فإن دراسة الحجم الأمثل للوحدة في مسح إحصائي واسع المدى هو مهمة ضخمة . وقدم Jessen (1942) مثالاً

جيدًا يتعلق بالمعـاينة الزراعية. ونعطي في الجدول (٩ ـ ٣) نبذة من نتائجه. وهي تقـارن 4 حجـوم للوحدة ـ قطاع ربعي، قطاع نصفي، قطاع، وشريط من الأرض يتألف من قطاعين متجاورين. والقطاع هو مساحة ميل مربع، يحوي في المتوسط، أقل من 4 مزارع بقليل. وفي هذه المقارنة حُدّدت التكاليف التالية: التكلفة الكلية للحقل

جدول (٩ ـ ٣) تقدير الأخطاء المعيارية (بالنسبة المئوية) من أجل أربعة حجوم من الوحدات، مع معاينة عشوائية بسيطة

المفردات	الوحدة الأفضل	$2S$	S	$S/2$	$S/4$
عدد الخنازير	$S/2$	6.2	5.3	4.9	5.0
عدد الخيول	$S/2$	4.2	3.6	3.3	3.4
عدد الأغنام	$2S$	14.3	14.9	15.7	17.4
عدد الفراريج	$S/4$, $S/2$	3.8	3.3	3.0	3.0
عدد البيض في اليوم السابق	$2S$	4.7	4.9	5.2	5.7
عدد القطيع	$S/2$	5.5	4.8	4.6	4.7
عدد البقرات التي حُلبت	$S/2$	4.4	3.8	3.6	3.7
عدد جالونات الحليب	$S/2$	4.9	4.4	4.2	4.4
محاصيل منتجات الحليب ومشتقاته	$S/2$	6.0	5.4	5.2	5.5
عدد الفدادين القابلة للزراعة	$S/2$	3.5	3.0	2.8	2.9
عدد فدادين الذرة	$S/2$	4.4	3.8	3.5	3.7
عدد فدادين الشوفان	$S/4$	7.0	5.6	4.8	4.6
إنتاج الذرة	$S/4$	2.5	2.0	1.7	1.6
إنتاج الشوفان	$S/2$	1.8	1.6	1.5	1.6
النفقات التجارية للإطعام	$S/4$	21.8	16.7	13.6	12.6
(operator) النفقات الإجمالية	$S/4$	12.0	9.6	8.1	7.8
(operator) المحاصيل الكلية	$S/4$	9.8	7.7	6.5	6.2
(operator) الدخل النقدي الصافي	$S/4$	9.5	7.8	6.9	6.8

($ 1000) ، طول الاستبيان (يستغرق 60 دقيقة لإتمامه)، ونفقة السفر (خمس سنتات لكـل ميل)، لأن الـدقـة النسبية الصـافية تتغير عند تحوّل أي من هذه المتغيرات. والتكاليف محسوبة وفقًا لأسعار 1939 .

والبيانـات في الجـدول هي الأخطاء المعيارية النسبية (بالنسبة المئوية) لتقدير المتوسطات على أساس المزرعة الواحدة من أجل 18 مفردة. وليست هناك وحدة مفضّلة لجميع المفردات. والقطاعان الربعي والنصفي متفوقان، على أي حال، على الوحدات الأكـبر من أجـل جميع المفـردات باستثنـاء اثنتين منها، وهناك القليل مما نختاره بين القطاعين الربعي والنصفي. ومن المحتمل أن يكون القطاع النصفي مفضلا، لأن مسألة التعرف على الحدود بدقة تكون أسهل.

(٩ - ٣) مقارنات دقة جرت باستخدام معلومات مسح إحصائي

في مثـال المشاتل حصلنا على التباينات لأنواع مختلفة من الوحدات من تعداد كامل للمجتمع. وعلى أي حال فإنه، باستثناء حالة مجتمعات صغيرة، يندر أن يكون القيام بمسح إحصائي شامل لمجرّد إجراء مقارنات أمرًا عمليًا، وفي العادة كثيرًا ما تأتي المعلومات حول الوحدة المثلى كحصيلة جانبية ذكية لمسح عينة غايته الرئيسة هي القيام بتقديرات.

لنفرض أنه يمكن تقسيم كل وحدة إلى M من الوحدات الأصغر. وبدلاً من تسجيل المجـاميع لكـل وحـدة «كبيرة» فقط من وحدات العيّنة، نسجل المعلومات الإحصائية بصورة منفصلة في كل من الوحدات الصغيرة الـ M . فيمكن عندئذ القيام بمقارنة بين دقة الوحدات الكبيرة والصغيرة. وسنفترض في البداية عيّنة عشوائية بسيطة حجمها n .

ويمكن حساب تحليل التباين في الجدول (٩ - ٤) من العيّنة.

جدول (٩ ـ ٤) تحليل التباين لبيان العيّنة (على أساس الوحدة الصغرى)

	درجات الحرية	متوسط المربعات
ما بين الوحدات الكبرى	$(n-1)$	s_b^2
ما بين الوحدات الصغرى ضمن الوحدات الكبرى	$n(M-1)$	s_w^2
ما بين الوحدات الصغرى في العيّنة	$nM-1$	$s^2 = \frac{(n-1)s_b^2 + n(M-1)s_w^2}{nM-1}$

وتقدير تباين وحدة كبرى (على أساس الوحدة الصغرى) هو s_b^2. ويمكن التفكير بأن متوسط المربعات بين جميع الوحدات الصغرى في العيّنة يمكن أن يكون تقديرًا ملائمًا لتباين الوحدة الصغرى أي أن:

$$s^2 = \frac{(n-1)s_b{}^2 + n(M-1)s_w{}^2}{nM-1} \qquad (9.5)$$

وهـذا التقدير منحاز انحيازًا طفيفًا مع أنه، في الغالب، مُرضٍ، وذلك لأن العيّنة ليست عيّنة عشوائية بسيطة من الوحدات الصغرى، طالما أن هذه الوحدات قد جرت معاينتها من زُمَر متجاورة في كل منها M من الوحدات.

ونحصل على تقدير غير منحاز من العيّنة عن طريق القيام بتحليل التباين، كما في الجدول (٩ ـ ٥)، وذلك في كامل المجتمع، الذي يحوي N من الوحدات الكبرى و N M من الوحدات الصغرى.

جدول (٩ ـ ٥) تحليل التباين في كامل المجتمع (على أساس الوحدة الصغرى)

	درجات الحرية	متوسط المربعات
ما بين الوحدات الكبرى	$N-1$	S_b^2
ما بين الوحدات الصغرى ضمن الوحدة الكبرى	$N(M-1)$	S_w^2
ما بين الوحدات الصغرى في المجتمع	$NM-1$	$S^2 = \frac{(N-1)S_b^2 + N(M-1)S_w^2}{NM-1}$

ووفقـا لتعريفه، فإن تباين المجتمع بين الوحدات الصغرى معطى في السطر الأخير من الجدول، أي أن:

$$S^2 = \frac{(N-1)S_b{}^2 + N(M-1)S_w{}^2}{NM-1} \qquad (9.6)$$

ومع المعاينة العشوائية البسيطة، يكون s_b^2 في الجدول (٩ ـ ٤) تقديرًا غير منحاز لـ S_b^2، (وينتج هذا من الفقرة ٢ ـ ٣). ويمكن البرهان بسهولة على أن s_w^2 هو تقدير غير منحاز لِـ S_w^2. وبالتالي فإن:

$$\hat{S}^2 = \frac{(N-1)s_b{}^2 + N(M-1)s_w{}^2}{NM-1} \qquad (9.7)$$

هو تقدير غير منحاز للتباين S^2 بين جميع الوحدات الصغرى في المجتمع.

ومن الواضح أن هذه العبارة هي تقريبًا العبارة الأبسط نفسها،

$$\hat{S}^2 \doteq \frac{s_b{}^2 + (M-1)s_w{}^2}{M} \qquad (9.8)$$

وإذا كان $n>50$ فإن (9.5) الخاصة بِـ S^2 تختزل أيضًا إلى (9.8)، بحيث يكون s^2 تقريبًا مُرضيًا لِـ S^2 في حالة $n>50$.

والتقديران s_b^2 (في حالة الوحدة الكبيرة) و $\hat{S}^2$ (في حالة الوحدة الصغيرة) محسوبان وفقًا لأساس مشترك ويمكن تعويضهما في النظرية (٩-١)، نتيجة (٢).

وإذا كانت العيّنة كبيرة، فيمكن قياس الوحدات الصغرى في عيّنة جزئية عشوائية من الوحدات الكبرى (مثلاً 100 من أصل 600). أو على الوجه الآخر، يمكن قياس وحدتين صغيرتين نختارهما عشوائيًا من كل وحدة كبرى. ويمكن أن نتقصّى آنيًا أكثر من حجم واحد للوحدة الصغرى، شريطة أن نأخذ معلومات إحصائية تعطي تقديرًا غير منحاز لـ S_w^2 في كل وحدة صغرى.

وفي حال المعاينة الطبقية، يمكن، وفقًا لهذه الطرق، تقدير التباينين، في كل طبقة على حدة، من أجل الوحدتين الصغرى والكبرى، ثم نعوّض في العلاقة المناسبة الخاصة بتباين تقدير من عيّنة طبقية.

مثـال

المعلومـات الإحصـائية مستقـاة من عيّنـة زراعية مأخـوذة في نورث كارولاينا عام 1942 لتقدير العمالة في المزارع (Morgan, Finkner و Monroe 1943) كانت طريقة سحب العينة هي وضع نقاط على الخريطة، بصورة عشوائية، واختيار المزارع الثلاث الأقرب إلى كل نقطة كوحدات معاينة. ولا يوصى باستخدام هذه الطريقة، لأنِ فرصة اختيار المزرعة الكبيرة للعينة أكبر من فرصة اختيار المزرعة الصغيرة، ولمزرعة معزولة فرصة أكبر من مزرعة واقعة في منطقة مكتظة بالمزارع. وسنتجاهل هنا أية تأثيرات لمثل هذا الانحياز.

ومن بيان العيّنـة من مزارع بمفردها، يمكننا مقارنة مجموعة المزارع الثلاث، كوحدة معاينة، مع وحدة معاينة تضم مزرعة بمفردها، والمفردة التي اختيرت هي عدد العمال الذين يتقاضون أجرًا، وقد قُسمت العيّنة إلى طبقات، وكانت الطبقة زمرة من النواحي المتشابهة من حيث كثافة مجتمع المزارع فيها، ومن حيث نسبة الأرض المزروعة الى أراضي المزرعة. وبما أن كسر المعاينة 1.9% فيمكن تجاهل الـ ت م م.

$$V(\hat{Y}_{st}) = \sum_h \frac{N_h{}^2 S_h{}^2}{n_h}$$

والإجراء الصحيح هو أن نحسب ضمن كل طبقة على حدة $N_h{}^2S_h{}^2/n_h$، وذلك من أجل النوعين من الوحدات، مستخدمين تحليل التباين والعبارة (9.8). وسنستخدم إجراءً أبسط كتقريب.

وقد احتوت الطبقة، بصورة عامة، بين 300 و 450 مزرعة، وقد أُخذ من كل طبقـة وحـدتي معاينة أو ثلاث وحدات، حيث تتضمن كل وحدة 3 مزارع، بحيث تصبح المعاينة تناسبية تقريبًا. وبفرض أن التناسب محقق أي $n_h/N_h=n/N$، يمكن كتابة:

$$V(\hat{Y}_{st}) = \frac{N}{n}\sum N_h S_h{}^2 \doteq \frac{N^2}{n}\bar{S}_h{}^2$$

هذا إذا فرضنا أيضًا أن S_h^2 لا يتغير كثيرًا بين طبقة وأخرى، بحيث يمكن أن نضع بدلاً من كل منها متوسطها $\bar{S}_h{}^2$.

وقد تمّ الحصول على تقديرات لِـ $\bar{S}_h^2$ من تحليل التباين في الجدول (٩ - ٦) القائم على أساس المزرعة الواحدة.

جدول (٩ - ٦) تحليل تباين عيّنة (عدد العمال المأجورين) (على أساس المزرعة الواحدة)

	درجات الحرية	متوسط المربعات
ما بين الوحدات ضمن الطبقات	825	6.218
ما بين المزارع ضمن الوحدات	2768	2.918
ما بين المزارع ضمن الطبقات	3593	3.676

وفي حالة زمرة من ثلاث مزارع يخدم متوسط المربعات $\bar{s}_{h3}{}^2 = 6.218$ كتقدير لِـ $\bar{S}_h^2$ على أساس المزرعة الواحدة. وفي حالة مزرعة بمفردها، مستخدمين (9.8)، نجد:

$$\hat{\bar{S}}^2 = \frac{6.218 + 2(2.918)}{3} = 4.018$$

وباستخدام النظرية (٩-١)، نتيجة (٢)، يشير الرقمان 6.218 الموافق لزمرة المزارع الثلاث، و 4.018 الموافق لمزرعة بمفردها، إلى التباينات النسبية التي حصلنا عليها من أجل حجم كلي مثبت للعيّنة. وتعطي مجموعة المزارع حوالي ثلثي دقة مزرعة بمفردها. ويمكن لاعتبارات التكاليف أن نجعل النتيجة في صالح الوحدة التي تتضمن ثلاث مزارع.

(٩ - ٤) التباين بدلالة الارتباط ضمن العنقود

يُعبّر أحيانًا عن علاقات التباين بدلالة معامل الارتباط ρ بين العناصر الواقعة في العنقود نفسه، وقد استُخدم هذا الأسلوب سابقًا في المعاينة النمطية (فقرة ٨ - ٣).

لتكن y_{ij} القيمة الملحوظة للعنصر j ضمن الوحدة i، وليكن y_i مجموع الوحدة فنحتاج في المعاينة العنقودية إلى التمييز بين نوعين من المتوسطات: المتوسط لكل وحدة $\bar{Y} = \sum y_i / N$، والمتوسط لكل عنصر $\bar{\bar{Y}} = \sum y_i / NM = \bar{Y}/M$. والتباين بين العناصر هو:

$$S^2 = \frac{\sum_{i,j} (y_{ij} - \bar{\bar{Y}})^2}{NM - 1}$$

وقد عُرّف (فقرة ٨ ـ ٣) معامل الارتباط العنقودي الداخلي، ρ، على الشكل:

$$\rho = \frac{E(y_{ij} - \bar{\bar{Y}})(y_{ik} - \bar{\bar{Y}})}{E(y_{ij} - \bar{\bar{Y}})^2} = \frac{2\sum_{i}\sum_{j<k} (y_{ij} - \bar{\bar{Y}})(y_{ik} - \bar{\bar{Y}})}{(M-1)(NM-1)S^2} \quad (9.9)$$

وعدد الحدود، E ، في البسط (الحدود الجدائية) هو $NM(M\text{-}1)/2$ أما E في المقام فيساوي $(NM\text{-}1)S^2/NM$.

نظرية (٩-٢)

إذا سُحبت عيّنة عشوائية بسيطة من n عنقودًا، كل منها يحوي M عنصرًا، من الـ N عنقودًا في المجتمع. فعندئذ يكون متوسط العيّنة لكل عنصر $\bar{\bar{y}}$ تقديرًا غير منحاز لـ $\bar{\bar{Y}}$ بتباين يساوي،

$$V(\bar{\bar{y}}) = \frac{1-f}{n} \cdot \frac{NM-1}{M^2(N-1)} S^2[1 + (M-1)\rho]$$

$$\doteq \frac{1-f}{nM} S^2[1 + (M-1)\rho] \quad (9.10)$$

حيث ρ معامل الارتباط ضمن العنقود.

برهان

لنرمز بـ y_i لمجموع العنقود i و $\bar{y} = \sum^{n} y_i/n$ وبالاستناد إلى النظريتين (٢ ـ ١) و(٢ ـ ٢)، نجد أن $\bar{y}$ تقدير غير منحاز لـ $\bar{Y}$ بتباين،

$$V(\bar{y}) = \frac{(1-f)}{n} \frac{\sum (y_i - \bar{Y})^2}{N-1}$$

إلا إن $\bar{Y} = M\bar{\bar{Y}}$ و $\bar{y} = M\bar{\bar{y}}$. وبالتالي يكون $\bar{\bar{y}}$ تقديرًا غير منحاز لـ $\bar{\bar{Y}}$ بتباين يساوي،

$$V(\bar{\bar{y}}) = \frac{1-f}{nM^2} \frac{\sum (y_i - \bar{Y})^2}{N-1} \quad (9.11)$$

ولكن،

$$(y_i - \bar{Y}) = (y_{i1} - \bar{\bar{Y}}) + (y_{i2} - \bar{\bar{Y}}) + \cdots + (y_{iM} - \bar{\bar{Y}})$$

فإذا أخذنا المربعات وجمعنا فوق جميع العناقيد الـ N ، نجد،

$$\begin{aligned} \sum_i^N (y_i - \bar{Y})^2 &= \sum_i^N \sum_j^M (y_{ij} - \bar{\bar{Y}})^2 + 2 \sum_i^N \sum_{j<k}^M (y_{ij} - \bar{\bar{Y}})(y_{ik} - \bar{\bar{Y}}) \\ &= (NM-1)S^2 + (M-1)(NM-1)\rho S^2 \\ &= (NM-1)S^2[1+(M-1)\rho] \end{aligned} \quad (9.12)$$

مستخدمين تعريف ρ في (9.9) . لنعوّض $V(\bar{\bar{y}})$ في (9.11) . فهذا يعطي،

$$V(\bar{\bar{y}}) = \frac{1-f}{n} \cdot \frac{NM-1}{M^2(N-1)} S^2[1+(M-1)\rho]$$

وهو المطلوب .

وإذا أخذنا عيّنة عشوائية من nM عنصرًا، فإن العلاقة الموافقة لِـ $V(\bar{\bar{y}})$ هي العلاقة (9.10) نفسها باستثناء الحد بين قوسين . ويبيّن العامل :

$$1+(M-1)\rho$$

مدى تغير التباين عند استخدام عنقود كوحدة معاينة بدلاً من عنصر بمفرده . وهذا العمل هو إذًا أثر التصميم deff الذي عرّفه Kish لعناقيد حجمها M (فقرة ٤ ـ ١١) . وإذا كان $\rho>0$ يكون العنقود أقل دقة من أجل حجم معطى للعيّنة . وإذا كان $\rho<0$ ، كما يحدث أحيانًا، فإن العنقود يكون أكثر دقة . والنظرية (٩ ـ ٢) هي تعميم بسيط للنظرية (٨ ـ ٢) بالفصل الثامن .

ويمكن إعطاء عبارة بديلة لِـ ρ لنرمز بِـ S_b^2 للتباين بين مجاميع العناقيد، على أساس وحدة بمفردها . فعندئذ،

$$\sum (y_i - \bar{Y})^2 = (N-1)MS_b^2$$

ويمكن كتابة المعادلة (9.12) على الشكل،

$$(N-1)MS_b^2 = (NM-1)S^2[1+(M-1)\rho]$$

بحيث يكون،

$$\rho = \frac{(N-1)MS_b{}^2 - (NM-1)S^2}{(NM-1)(M-1)S^2} \doteq \frac{S_b{}^2 - S^2}{(M-1)S^2} \qquad (9.13)$$

عندما تكون الحدود في $1/N$ مهملة .

والجدير بالملاحظة هو قيمة متوسط مربعات ما ضمن العناقيد :

$$S_w{}^2 = \sum_i^N \sum_j^M (y_{ij} - \bar{y}_i)^2 / N(M-1) \qquad (9.14)$$

وبالاستناد إلى تحليل التباين في تصنيف أحادي نجد،

$$(NM-1)S^2 = \sum_i^N (y_i - \bar{Y})^2/M + \sum_i^N \sum_j^M (y_{ij} - \bar{y}_i)^2$$

$$= \frac{(NM-1)}{M} S^2[1 + (M-1)\rho] + N(M-1)S_w{}^2$$

وبالتالي، من (9.12) نجد،

$$S_w{}^2 = \frac{NM-1}{NM} S^2(1-\rho) \doteq S^2(1-\rho) \qquad (9.15)$$

وقـد قدّم Hansen ، Hurwitz ، و Madow (1953) ، مناقشة جيدة للقيم العـددية لِـ ρ من أجـل مفـردات مختلفـة وحجـوم مختلفـة للعنقود، وهم يعتبرون ρ «كمقياس تجانس» للعنقود .

(٩ ـ ٥) دوالّ تبايـن

في بعض أنـواع المسوح الإحصائية، على سبيل المثال، معاينة تربة، أو جني محصول، أو مسوح إحصائية خاصة بالمزارع حيث تُستخدم وحدات معاينة مساحيّة، يمكن أن يشكل مشكلة حجم الوحدة العنقودية متغيراً مستمرًا تقريبًا . وعند البحث عن أفضل وحدة، لا تكون المشكلة مشكلة اختيار بين حجمين أو ثلاثة حجوم محددة تمّت تجربتها، وإنما إيجاد القيمة المثلى لِـ M باعتبارها متغيّراً مستمرًّا . وتستدعي هذه المسألة طريقة للتنبؤ بالتباين S_b^2 بين وحدات المجتمع كدالّة في M . وبالاستفادة من تحليل التباين يمكن إيجاد S_b^2 إذا علمنا (ا) التباين S^2 بين جميع عناصر المجتمع و(ب)

التبـاين S_w^2 بين العنـاصر الـواقعـة في الـوحدة نفسها. وأسلوبنا هو التنبؤ بـ S_w^2 و S^2 وإيجاد S_b^2 بوساطة تحليل التباين.

ويُنتج بيان العيّنة تقديرين لِـ S^2 و S_w^2 من أجل حجم الوحدة المستخدمة فعلاً، وبما أن S^2 هو التباين بين العناصر، فإنه لا يتأثر بحجم الوحدة. ولكن S_w^2 ستتأثر. ويمكننا توقع ازديادها عندما يزداد حجم الوحدة الكبرى. وإذا كانت الوحدات الكبرى التي سندرسها تختلف في حجومها اختلافًا بسيطًا عن الوحدة المستخدمة فعلاً، فيمكن كتقريب أول اعتبار S_w^2 ثابتًا، مستخدمين التقدير الناتج عن بيان العيّنة. ويقترح البحث الذي قام به McVay (1947) أن هذا التقريب يمكن أن يكون في الغالب مُرضيًا.

وكتقريب أفضل، تمت محاولات (Jessen, 1942; Mahalanobis, 1944; و Hendrick, 1944) لتطوير قانون عام يتنبأ بكيفية تبدّل S_w^2 مع حجم الوحدة. وفي عدة مسوح إحصائية زراعية، بدا S_w^2 وكأنه يرتبط بِـ M وفقًا للعلاقة التجريبية،

$$S_w{}^2 = AM^g \qquad (g > 0) \tag{9.16}$$

حيث A و g ثابتان لا يعتمدان على M. وفي هذه العلاقة يتزايد S_w^2 باطراد كلما ازداد M. ويكون g عادة صغيراً. ويمكن توقع منحى من هذا النوع، عند وجود قوى تمارس نفوذًا مشابهًا على عناصر قريبة من بعضها. فالمناخ، ونوع التربة والطبوغرافيا، والوصول إلى الأسواق، تميل جميعها إلى أن تجعل للمزارع المتجاورة مقوّمات متشابهة.

ومن الناحية النظرية، فإن الصيغة مفتوحة للاعتراض، باعتبار أنها تجعل $S_w{}^2$ يتزايد بدون حدود عندما يتزايد M. وإذا فرضنا، وهو فرض يبدو منطقيًا، أنه لا يوجد أي ارتباط بين العناصر البعيدة عن بعضها، فقد تكون الصيغة الأكثر ملاءمة، هي صيغة يقترب فيها $S_w{}^2$ من حد أعلى في حالة M كبير. وعلى أي حال، فإن أي صيغة ستكون كافية إذا أعطت تلاؤما جيدًا فوق مدى M الذي تتطرق إليه الدراسة.

وفي حال ملاءمـة هذه الصيغة للواقع المدروس فإن الرسم البياني لِـ $\log S_w^2$ مقابل $\log M$ ينبغي أن يكون خطًا مستقيمًا. ولتقدير الثابتين $\log A$ و g نحتاج إلى قيم S_w^2 الموافقة لقيمتين لِـ M على الأقل. وعلى الأقل فإن ثلاث قيم لِـ M ضرورية من أجل أي تقويم لخطية الملاءمة.

ونجد من تحليل التباين في الجدول (٩ـ٥) أن،

$$S_b^2 = \frac{(NM-1)S^2 - N(M-1)S_w^2}{N-1}$$

$$= \frac{(NM-1)S^2 - N(M-1)AM^g}{N-1}$$

$$\doteq MS^2 - (M-1)AM^g \qquad (9.17)$$

وقد أشار Hendricks (1944) إلى أنه يمكن اعتبار المجتمع بكامله كوحدة معاينة كبرى واحدة تتضمن NM عنصرًا. وإذا صحّت (9.16) فعندئذ $S^2 = A(NM)^g$. وفائدة هذا التدبير هو أنه يمكن الآن تقدير قيم A و g من البيان الإحصائي لمسح استُخدمت فيه قيمة واحدة لِـ M فقط. والمعادلتان اللتان تقودان إلى التقديرين هما،

$$\log S_w^2 = \log A + g \log M \qquad (9.18)$$

$$\log S^2 = \log A + g \log (NM) \qquad (9.19)$$

وبالاستناد إلى (9.17) تصبح صيغة S_b^2 على الشكل:

$$S_b^2 \doteq AM^g[MN^g - (M-1)] \qquad (9.20)$$

ولا تجهّزنا هذه الطريقة بإمكانية تدقيق صحة العلاقة (9.16) وقد تصحّ العلاقة، وبجودة كافية، في حالة قيم صغيرة لِـ M، ولكنها تفشل في حالة قيم كبيرة في مثل الحجم NM.

وقـد قُدّمت العـلاقة (9.16) كمثال على الطرائقية أكثر من كونها قانونًا عامًا. والقارىء الذي يواجه مسألة مشابهة، ينبغي له أن يضع ويختبر أي نوع من الصيغ التي تبدو أكثر ملاءمة لمادته. وفي بعض الحالات يمكن أن يكون $\log S_b^2$ دالة خطية في M كما اقترح Fairfield Smith (1938) في حالة بيان إحصائي يتعلق بالإنتاج.

(٩ ـ ٦) دالــة تكلفــة

في مسح إحصائي واسع المدى، تحتل طبيعة تكاليف العمل الميداني مكانًا كبيرًا في تحديد الوحدة المثلى. وكتوضيح لدور عوامل التكلفة، سنستعرض حالة تكلفة طوّرها Jessen (1942) لمسوح إحصائية تتعلق بالمزارع، كانت وحدتها الكبرى عناقيد من المزارع المتجاورة.

ونميز هنا مركبتين لتكلفة العمل الميداني، المركبة الأولى C_1MN ؟ تكلفة المقابلة وتكلفة السفر من مزرعة إلى مزرعة ضمن العنقود.

أما المركبة الثانية، $c_2\sqrt{n}$ ، فتقيس تكلفة السفر من عنقود إلى آخر. وتبيّن الاختبارات على خريطة أن هذه التكلفة تتغير، في حالة مجتمع مثبت، وفقًا للجذر التربيعي لعدد العناقيد تقريبًا. والتكلفة الكلية للعمل الميداني هي إذن،

$$C = c_1Mn + c_2\sqrt{n} \tag{9.21}$$

وبفرض أن المعاينة عشوائية بسيطة، ومع تجاهل عامل الـ ت م م فإن تباين المتوسط لكل عنصر $\bar{\bar{y}}$ هو S_b^2/nM وهذا يساوي بالاستناد إلى (9.17) ،

$$V(\bar{\bar{y}}) = \frac{S^2-(M-1)AM^{g-1}}{n} \tag{9.22}$$

ولتحديد الحجم الأمثل للوحدة، نجد M ، وبالمناسبة n أيضًا، التي تجعل V في نهايته الصغرى من أجل C مثبتة. والحل العام معقّد، مع أن تطبيقه في مسألة عددية ما لا يقدّم صعوبة كبيرة.

وبحسابات بسيطة يمكن الحصول على المعادلة التي تعطي القيمة المثلى لِـ M . إذ نحلّ أولاً معادلة التكلفة (9.21) كمعادلة من الدرجة الثانية في $\sqrt{n}$ ، وهذا يعطي،

$$\frac{2c_1M\sqrt{n}}{c_2} = \left(1+\frac{4Cc_1M}{{c_2}^2}\right)^{1/2} - 1 \tag{9.23}$$

والمعادلة المستخدمة لحساب القيمة الصغرى

$$C+\lambda V = c_1Mn + c_2\sqrt{n} + \lambda V$$

وبالتفاضل وملاحظة أن $\partial V/\partial n = -V/n$ ، نحصل على المعادلات،

$$n: \qquad c_1M+\tfrac{1}{2}c_2n^{-1/2} = -\frac{\lambda\,\partial V}{\partial n} = \frac{\lambda V}{n} \tag{9.24}$$

$$M: \qquad c_1n = -\frac{\lambda\,\partial V}{\partial M} \tag{9.25}$$

لنقسم (9.25) على (9.24) بحيث نحذف λ ، فيقودنا ذلك إلى،

$$\frac{n}{V}\frac{\partial V}{\partial M} = -\frac{c_1n}{c_1M+\frac{1}{2}c_2n^{-1/2}}$$

أو،

$$\frac{M}{V}\frac{\partial V}{\partial M} = -\frac{1}{1+c_2/2c_1M\sqrt{n}} \tag{9.26}$$

وإذا عوضنا $\sqrt{n}$ من (9.23) ، نحصل بعد التبسيط على،

$$\frac{M}{V}\frac{\partial V}{\partial M} = \left(1+\frac{4Cc_1M}{c_2^{\,2}}\right)^{-1/2} - 1 \tag{9.27}$$

وبكتابة الطرف الأيسر من هذه المعادلة بكامله وتغيير إشارات الطرفين نجد،

$$\frac{AM^{g-1}[gM-(g-1)]}{S^2-(M-1)AM^{g-1}} = 1-\left(1+\frac{4Cc_1M}{c_2^{\,2}}\right)^{-1/2} \tag{9.28}$$

وهذه المعادلة تعطي القيمة لِـ M . ولا يتضمن الطرف الأيسر أيًّا من عوامل التكلفة، باعتبارها تعتمد فقط على شكل دالة تباين ونرى بوضوح أن كلا الطرفين دالة متزايدة في M ، من أجل $M \geq 1, g>0$ ، وذلك ضمن منطقة الدراسة . ولنفرض الآن أننا وجدنا حلاً من أجل قيم محددة لِـ C ، C_1 و C_2 ، ونرغب في دراسة تأثير الزيادة في C_1 على هذا الحل . فنلاحظ عندئذ أن الطرف الأيسر لا يعتمد على C_1 ، إلا أن الطرف الأيمن يزداد عندما تزداد C_1 . وعلى أي حال، فقد وُجد أن القيمة المثلى لِـ M تتناقص بسبب وجود C_1M في الطرف الأيمن . ويُنتج التناقص في C_2 تأثيراً مشابهاً .

والآن يتزايد C_1 إذا ازداد طول المقابلة، بينما يتناقص C_2 إذا أصبح السفر أرخص، أو إذا أصبحت المزارع في منطقة معينة أكثر كثافة . وهذه الحقائق تقود إلى

الاستنتاج بأن الحجم الأمثل للوحدة يصبح أصغر عندما،

يزداد طول المقابلة،

يصبح السفر أرخص،

تصبح العناصر (المزارع) أكثر كثافة،

تزداد الكمية الإجمالية للنقود المستخدمة (C) .

وهذه النتيجة مستقاة من النوع المستخدم لدالة التكلفة وستحتاج إلى إعادة نظر في حالة دالة مختلفة. إنها توضح حقيقة أن الوحدة المثلى ليست خاصة مميزة ثابتة للمجتمع، ولكنها تعتمد أيضًا على نوع المسح الإحصائي وعلى مستويات الأسعار والأجور.

ويعطي Hansen ، Hurwitz ، و Madow (1953) مناقشة ممتازة لبناء دوال تكلفة من أجل مسوح إحصائية تتضمن معاينة عنقودية.

(٩ - ٧) المعاينة العنقودية في حالة النسب

تنطبق الطرق نفسها على معاينة عنقودية في حالة النسب. فلنفرض أنه يمكن تصنيف العناصر الـ M في أي عنقود إلى صفين، وأن $p_i=a_i/M$ هي نسبة العناصر من C في العنقود i . تؤخذ عيّنة عشوائية بسيطة من n عنقودًا، ونستخدم المتوسط p للنسب الملحوظة p_i في العيّنة كتقدير لنسبة المجتمع p .

وكما نذكر (فقرة ٣-١٢) فإننا لا نستطيع استخدام نظرية ثنائية الحد لإيجاد $V(p)$ ، ولكن يجب تطبيق العلاقة الخاصة بالمتغيرات المستمرة على النسب p_i . وهذا يعطي،

$$V(p)=\frac{N-n}{Nn}\frac{\sum_{i=1}^{N}(p_i-P)^2}{N-1}\doteq\frac{N-n}{N^2n}\sum(p_i-P)^2 \tag{9.29}$$

وعلى الوجه الآخر، إذا أخذنا عيّنة عشوائية بسيطة تحوي nM من العناصر، فاستنادًا إلى نظرية ذي الحدين (نظرية ٣-٢) نحصل على تباين p وفق الصيغة:

$$V_{bin}(p)=\frac{(NM-nM)}{NM-1}\frac{PQ}{nM}\doteq\frac{N-n}{N}\frac{PQ}{nM} \tag{9.30}$$

وذلك إذا كان N كبيراً . وبالتالي يبيّن عامل أثر التصميم deff ، وهو:

$$\frac{V(p)}{V_{bin}(p)} \doteq \frac{M\sum(p_i-P)^2}{NPQ} \quad (N \text{ كبيرا}) \tag{9.31}$$

التغيّر النسبي في التباين الذي يعود إلى استخدام العناقيد . والقيم العددية لهذا العمل مفيدة في صنع تقديرات تمهيدية لحجم العيّنة في حالة معاينة عنقودية . ويُقدّر أولاً الحجم المطلوب للعيّنة بالاستناد إلى علاقة التوزيع الثنائي، ثم يُضرب بالعامل للحصول على الحجم الذي سيكون ضرورياً في حالة معاينة عنقودية . ولتوضيح الفكرة انظر Cornfield (1951) .

وإذا كانت حجوم العناقيد M_i متغيرة فإن التقدير $p=\sum a_i/\sum M_i$ هو التقدير النسبة . وتباينه معطى تقريبًا بالعلاقة (فقرة ٣-١٢)،

$$V(p) \doteq \frac{N-n}{Nn\bar{M}^2}\,\frac{\sum_{i=1}^{N} M_i^2(p_i-P)^2}{N-1} \tag{9.32}$$

حيث $\bar{M}=\sum M_i/N$ هو الحجم المتوسط للعنقود .

وإذا قورنت هذه العيّنة بعيّنة عشوائية بسيطة من $n\bar{M}$ من العناصر، فإننا نجد كتعميم لِـ (9.31) ،

$$\frac{V(p)}{V_{bin}(p)} \doteq \frac{\sum M_i^2(p_i-P)^2}{N\bar{M}PQ} \tag{9.33}$$

وكما في حالة المتغيرات المستمرة، فإنه يمكن تقصّي نسبة حجم العنقود إلى تباين ما بين العناقيد، إما بالتعبير عن العامل في المعادلتين (9.31) و (9.33) كدالّة في $\bar{M}$ أو بالاتجاه إلى علاقة بين تباين ما ضمن العنقود و $\bar{M}$. وإذا خصصنا القيمة 1 إلى أي وحدة تقع ضمن الصف C و 0 إلى أي وحدة أخرى، تكون المعادلة الأساسية لتحليل التباين من أجل M مثبت هي :

$$NMP(1-P) = M\sum(p_i-P)^2 + M\sum p_i(1-p_i) \tag{9.34}$$

مجموع المربعات الكلي = مجموع مربعات ما بين العناقيد + مجموع مربعات ما ضمن العناقيد .

ويمكننا أن نحسب من هذه العلاقة متوسط مربعات ما ضمن العناقيد ونرسم خطها البياني كدالة في M . ويصف Mc Vay (1947) كيف يمكن استخدام هذا التحليل لتقصّي الحجم الأمثل للعنقود.

تمارين

(٩ ـ ١) من أجل المعلومات الإحصائية في الجدول (٩ـ١)، قارن الدقة النسبية الصافية للأنواع الأربعة من الوحدات، وذلك عندما يكون الهدف هو تقدير العدد الكلي للأغراس في المسكبة بخطأ معياري يساوي 200 غرسة (لاحظ أن عامل الـ ت م م مأخوذ بعين الاعتبار).

(٩ ـ ٢) من أجل المعلومات الإحصائية في الجدول (٣ـ٥) في الفصل الثالث، قدّر الدقة النسبية للأسرة مقابل الفرد، وذلك لتقدير نسبة الجنس ونسبة الأفراد الذين راجعوا طبيبًا في الأشهر الإثني عشر السابقة، مفترضين أن المعاينة عشوائية بسيطة.

(٩ ـ ٣) قسمنا مجتمعًا من 2500 من العناصر إلى عشر طبقات، تحوي كل منها 5 وحدات كبرى تتألف كل منها من 50 عنصرًا. وتحليل تباين المجتمع، على أساس العنصر الواحد، ولمفردة واحدة، هو كما يلي:

	درجات الحرية	متوسط المربعات
ما بين الطبقات	9	30.6
مابين الوحدات الكبرى ضمن الطبقات	490	3.0
مابين العناصر ضمن الوحدات الكبرى	2000	1.6

متجاهلين عامل الـ ت م م، هل تكون الدقة النسبية للوحدة الكبيرة إلى الوحدة الصغيرة أكبر في حالة المعاينة العشوائية البسيطة منها في المعاينة العشوائية الطبقية (محاصّة تناسبية)؟

(٩ ـ ٤) قسمنا مجتمعًا يحتوي على $L\bar{N}M$ من العناصر إلى L من الطبقات، تتألف كل منها من $\bar{N}$ وحدة كبيرة، وكل من هذه الوحدات الكبيرة تحتوي على M من الوحدات الصغيرة. وتأتي الكميات التالية من تحليل تباين المجتمع على أساس العنصر الواحد:

S_1^2 = متوسط مربعات ما بين الطبقات

S_2^2 = متوسط مربعات ما بين الوحدات الكبيرة ضمن الطبقات

S_3^2 = متوسط مربعات ما بين العناصر ضمن الطبقات .

إذا كان $\bar{N}$ كبيراً وتجاهلنا عامل الـ ت م م ، فبيّن أن الدقة النسبية للوحدة الكبيرة إلى الوحدة الصغيرة (عنصر) تتحسن من خلال التقسيم إلى طبقات إذا كان :

$$\frac{(M-1)}{S_1^2} < \frac{M}{S_2^2} - \frac{1}{S_3^2}$$

(٩ ـ ٥) في مسح إحصائي في الريف كانت وحدة المعاينة عنقودًا من المزارع ، ووجدنا أن كلفة أخذ عيّنة تتضمن n وحدة هي ،

$$C = 4tMn + 60\sqrt{n}$$

حيث t هو الزمن بالساعات الذي نقضيه في الحصول على الأجوبة من مزارع واحد . وإذا أنفقنا 2000 $ على هذا المسح ، نجد أن قيم n في حالة $M=1,2,10$ ؛ $t=\frac{1}{2},2$ هي كما يلي :

	M		
	1	5	10
$t = \frac{1}{2}$ hr	400	131	74
$t = 2$ hr	156	40	21

تحقق من قيمتين من هذه القيم للتأكد من أنك تفهم استخدام العلاقة .

وتباين متوسط العيّنة (متجاهلين الـ ت م م) هو،

$$\frac{S^2}{Mn}[1+(M-1)\rho]$$

وإذا كان $\rho=0.1$ من أجل جميع قيم M بين 1 و 10 ، فما هي حجم الوحدة الأكثر دقة في حالة (ا) t = نصف ساعة ، (ب) $t=2$ ساعة؟ كيف تفسر الفرق بين النتيجتين؟

(٩ ـ ٦) إذا توافر 5000 $ لمسح إحصائي ، فهل تتوقع تناقص أو تزايد الحجم الأمثـل للوحـدة (بالمقارنة مع 2000 $)؟ أعط أسبابًا. ويمكنك ، إذا رغبت ، إيجاد الحجم الأمثل للتحق من صحة محاكمتك .

الفصل التاسع (أ)

المعاينة العنقوديّة وحيدة المرحلة: عناقيد ذات حجوم غير متساوية

(١٩ ـ ١) وحدات عنقودية ذات حجوم غير متساوية

في معظم التطبيقات تتضمن الوحدات العنقودية (مثلاً، مناطق، مدن، جادّات مدينـة) أعـدادًا مختلفـة من العناصر أو الوحدات الجزئية (وحدات مساحية، أُسر، أشخاص). ويعالج هذا الفصل بعضًا من الطرق العديدة التي تمّ الوصول إليها سواء أكانت طرقًا لاختيار العيّنة أو للتقدير، وذلك في حالة وحدات عنقودية بحجوم غير متساوية. ليكن M_i عدد العناصر في الوحدة i . فنعلم سابقًا طريقتين مألوفتين لتقدير مجموع المجتمع Y للقياسات y_{ij} .

عيّنة عشوائية بسيطة من العناقيد: تقدير غير منحاز

كما سبق، لنرمز بِـ $y_i = \sum_{i=1}^{M_i} y_{ij} = M_i \bar{y}_i$

لمجموع المفردة في الوحدة العنقودية i . فإذا فرضنا عيّنة عشوائية بسيطة حجمها n من الوحدات الـ N في المجتمع، فإن تقديرًا غير منحاز لِـ Y هو (استنادًا إلى النظرية ٢ـ١ نتيجة)،

$$\hat{Y} = \frac{N}{n} \sum_{i=1}^{n} y_i \tag{9A.1}$$

وبالاستناد إلى النظرية (٢ـ٢)، يكون تباينه.

$$V(\hat{Y}) = \frac{N^2(1-f)}{n} \frac{\sum_{i=1}^{N} (y_i - \bar{Y})^2}{N-1} \qquad (9A.2)$$

حيث $\bar{Y} = Y/N$ هو متوسط المجتمع على أساس الوحدة العنقودية.

وغالبًا ما وُجد التقدير $\hat{Y}$ فقير الدقة. ويحدث هذا عندما تختلف الـ $\bar{y}_i$ (المتوسطات على أساس العنصر الواحد) اختلافًا بسيطًا من وحدة إلى وحدة بينما يتغير M_i كثيرًا. وفي هذه الحالة يتغير $y_i = M_i \bar{y}_i$ كثيرًا أيضًا من وحدة إلى وحدة ويكون التباين في (9A.2) كبيرًا.

عيّنة عشوائية بسيطة من العناقيد: تقدير النسبة إلى الحجم

ليكن،

$$M_0 = \sum_{i=1}^{N} M_i = \text{العدد الكلي للعناصر في المجتمع}$$

وإذا كانت المقادير M_i، وبالتالي M_0، جميعها معروفة، نجد تقديرًا بديلاً هو التقدير النسبة حيث نتخذ M_i كمتغير مساعد x_i.

$$\hat{Y}_R = M_0 \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} M_i} = M_0 \; = \text{(متوسط العيّنة لكل عنصر)}$$

ووفق رموز التقدير النسبة نجد أن نسبة المجتمع $R = Y/X = Y/M_0 = \bar{\bar{Y}}$، وهو متوسط المجتمع لكل عنصر. وبالاستناد إلى النظرية (٦-١)، مفترضين أن عدد العناقيد في العيّنة كبير، نجد

$$V(\hat{Y}_R) \doteq \frac{N^2(1-f)}{n} \frac{\sum_{i=1}^{N} (y_i - M_i \bar{\bar{Y}})^2}{N-1} \qquad (9A.3)$$

$$\doteq \frac{N^2(1-f)}{n} \frac{\sum^{N} M_i^2 (\bar{y}_i - \bar{\bar{Y}})^2}{N-1} \qquad (9A.4)$$

وكما يظهر من (9A.4) ، فإن تباين $\hat{Y}_R$ يعتمد على التغير بين المتوسطات لكل عنصر وقد وُجد أنه على الغالب أصغر بكثير من $V(\hat{Y})$.

نلاحظ أن $\hat{Y}_R$ يستدعي معرفة المجموع M_0 لجميع العناصر M_i ، بينما لا يستدعي $\hat{y}$ ذلك. والعكس صحيح عندما نكون في صدد تقدير متوسط المجتمع لكل عنصر. وفي هذه الحالة تكون التقديرات الموافقة

$$\text{متوسط العيّنة لكل عنصر} = \hat{\bar{Y}} = \frac{\hat{Y}}{M_0} = \frac{N}{nM_0}\sum^{n} y_i,\quad \hat{\bar{Y}}_R = \frac{\hat{Y}_R}{M_0} = \frac{\sum^{n} y_i}{\sum^{n} M_i}$$

وهكذا لا يتطلّب $\hat{\bar{Y}}_R$ إلا معرفة الـ M_i التي تقع في العيّنة التي اخترناها.

(١٩ ـ ٢) المعاينة في حالة احتمال متناسب مع الحجم

إذا كانت المقادير M_i جميعها معروفة، فتوجد طريقة أخرى طوّرها Hansen و Hurwitz (1943) وهي اختيار الوحدات باحتمالات متناسبة مع حجومها M_i . ونوضح طريقة اختيار وحدة بمفردها من المجتمع الصغير التالي المؤلف من $N=7$ وحدات.

الوحدة	الحجم M_i	$\sum M_i$	المدى المخصّص
1	3	3	1–3
2	1	4	4
3	11	15	5–15
4	6	21	16–21
5	4	25	22–25
6	2	27	26–27
7	3	30	28–30

وقد شُكّلت المجاميع التجميعية لِـ M_i . ولاختيار وحدة نسحب عددًا عشوائيًا بين 1 و $M_0=30$. لنفرض أن هذا العدد 19 . ففي العمود $\sum M_i$ يقع العدد 19 في الوحدة الرابعة التي تغطى المدى من 16 إلى 21 بما فيه الطرفان (16 و 21) ومع هذه الطريقة في السحب يكون احتمال اختيار أي وحدة أخرى متناسبًا مع حجمها.

وتكون هذه الطريقة في اختيار وحدة مريحة، فقط عندما يكون N معتدلاً، أو في المعـاينة الطبقية عندما تكون الحجوم N_h معتدلة أو صغيرة، ذلك لأن تجميع المقادير M_i يمكن أن يستهلك الكثير من الوقت في حالة N كبير (مثلاً $N=20,000$). وفي هذه الحالة، أعطى Lahiri (1951) طريقة بديلة تتجنب عملية التجميع. ليكن M_{max} أكبر المقادير M_i ولنسحب عددًا عشوائيًا بين 1 و N؛ لنفرض أن هذا العدد هو i. نسحب الآن عددًا عشوائيًا آخر m بين 1 و M_{max}. فإذا كان m أقل من أو يساوي M_i، نختار الوحدة i. وإذا لم يكن، نسحب زوجًا آخر من الأعداد العشوائية. ومن الطبيعي أن تتعرض هذه الطريقة لأقل ما يمكن من الاعتراضات عندما لا تختلف المقادير كثيراً بعضها عن بعض.

لنفرض الآن أن $n>1$. ولنفترض مؤقتًا أن المعاينة مع الإعادة. فلاختيار وحدة ثانية وفقًا لطريقة التجميع نسحب عددًا عشوائيًا جديدًا بين 1 و 30. إلا أنه، وخلافًا للمعاينة بدون إعادة، لا نحظّر اختيار الوحدة 4 للمرة الثانية. ومع هذه القاعدة، تبقى احتمالات الاختيار متناسبة مع الحجوم عند كل سحب وميزة الاختيار مع الإعادة هو بساطة العلاقات الخاصة بالتباينات الصحيحة أو المقدّرة للتقديرات.

وفي حالـة المعـاينـة بدون إعـادة، على الـوجـه الآخر (فقرة ١٩ ـ ٦)، يكون الاحتفاظ باحتمالات اختيار متناسبة مع الحجوم المختارة أكثر صعوبة، ويصبح عاجلاً أم آجلاً نوعًا من المستحيل مع تزايد n. ويمكن رؤية هذا في الحالة المتطرفة $n=7$ في المثال السابق (مع أنها غير عملية). وإذا قمنا بالاختيار بدون إعادة، فقد يكون من المؤكـد اختيار كل وحـدة من وحدات المجتمع وذلك بصرف النظر عن الحجوم الأصلية M_i. وعـلى أي حال، ففي حالـة معاينة طبقية تكون الحجوم N_h فيها صغيرة، تمّت أبحـاث كثـيرة (فقـرة ١٩ ـ ٦) لتطوير طرق عملية للمعاينة باحتمالات غير متساوية وبدون إعادة.

(١٩ - ٣) الاختيار باحتمالات غير متساوية مع الإعادة

لنختبر الوحدة i باحتمال M_i/M_0 ومع الإعادة، حيث $M_0=\sum M_i$ فسنبيّن أن تقديرًا غير منحاز لمجموع المجتمع Y هو،

$$\hat{Y}=\frac{M_0}{n}(\bar{y}_1+\bar{y}_2+\cdots+\bar{y}_n) \qquad (9A.5)$$

(متوسط متوسطات الوحدات لكل عنصر) $M_0 =$

وللمقارنة مع طرق لاحقة سنرمز لهذا التقدير بـ $\hat{Y}_{pps}$. وبالإضافة إلى ذلك،

$$V(\hat{Y}_{pps})=\frac{M_0}{n}\sum_{i=1}^{N} M_i(\bar{y}_i-\bar{\bar{Y}})^2 \qquad (9A.6)$$

بحيث يعتمد تباين $\hat{Y}_{pps}$ مثله مثل تباين $\hat{Y}_R$ على تغير متوسطات الوحدات لكل عنصر.

وفي بعض التطبيقات نعلم الحجوم M_i بصورة تقريبية فقط. وفي تطبيقات أخرى لا يكون «الحجم» عدد العناصر في الوحدة ولكنه قياس لكبرها، يُعتقد أنه مرتبط ارتباطًا عاليًا مع مجموع الوحدة y_i. وعلى سبيل المثال، يمكن قياس «حجم» مستشفى بالعدد الإجمالي للأسرّة أو بالعدد المتوسط للأسرّة المشغولة فوق فترة زمنية ما. وبصورة مماثلة، يمكن استنباط مقاييس مختلفة لِـ«حجم» مطعم، مصرف، أو منطقة زراعية. وبالتالي فإننا سنعتبر قياس حجم M'_i واحتمال الاختيار الموافق $z_i=M'_i/M'_0$ ، حيث $M'_0=\sum M'_i$. وإلى الحد الذي يتعلق بالنتائج النظرية يمكن أن تكون المقادير z_i أيّ مجموعة من الأعداد الموجبة مجموعها فوق المجتمع بكامله يساوي الواحد وسنبيّن أن،

$$\hat{Y}_{ppz}=\frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{z_i} \qquad (9A.7)$$

هو تقدير غير منحاز لِـ Y بتباين،

$$V(\hat{Y}_{ppz})=\frac{1}{n}\sum_{i=1}^{N} z_i\left(\frac{y_i}{z_i}-Y\right)^2 \qquad (9A.8)$$

وتَستخدم البراهين طريقة قدمناها في الفقرة (٢-١٠). ليكن t_i عدد المرات التي تظهر فيها الوحدة i في عيّنة محددة حجمها n ، حيث يمكن أن يأخذ t_i أيًّا من

القيم $1,2,\ldots,n$ ، لنأخذ توزيع التكرار المشترك للمقادير t_i من أجل جميع الوحدات الـ N في المجتمع .

وطـريقـة سحب عيّنة مكافئة لمسألة الاحتمال المعروفة التي نقذف فيها n كرة إلى N صنـدوقًـا الـخ . وعند كل قذفة يمثل z_i احتمال أن تذهب كرة إلى الصندوق i . وبالتالي يكون التوزيع المشترك للمقادير t_i هو توزيع متعدد الحدود،

$$\frac{n!}{t_1!t_2!\ldots t_N!} z_1^{t_1} z_2^{t_2} \ldots z_N^{t_N}$$

والخواص التالية لتوزيع المتغيرات t_i ، وهو التوزيع متعدد الحدود، هي خواص معروفة جيدًا .

$$E(t_i) = nz_i, \qquad V(t_i) = nz_i(1-z_i), \qquad \text{Cov}\,(t_i t_j) = -nz_i z_j \tag{9A.9}$$

نظرية (١٩-١)

إذا سحبنا عيّنة من n من الوحدات باحتمالات z_i ومع الإعادة، فعندئذ يكون

$$\hat{Y}_{ppz} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{z_i} \tag{9A.10}$$

تقديرًا غير منحاز لِـ Y بتباين،

$$V(\hat{Y}_{ppz}) = \frac{1}{n} \sum_{i=1}^{N} z_i \left(\frac{y_i}{z_i} - Y\right)^2 \tag{9A.11}$$

برهـان

يمكن كتابة

$$\hat{Y}_{ppz} = \frac{1}{n}\left(t_1 \frac{y_1}{z_1} + t_2 \frac{y_2}{z_2} + \cdots + t_N \frac{y_N}{z_N}\right) = \frac{1}{n} \sum_{i=1}^{N} t_i \frac{y_i}{z_i}$$

حيث يمتـد المجمـوع فوق جميع الـوحـدات في المجتمـع . وعند تكرار المعاينة تمثل المقـادير t المتغـيرات العشـوائية، بينما الـ y_i والـ z_i مجموعة من الأعداد المثبتة . ومنه، وباعتبار $E(t_i)=nz_i$ ، نجد من (9A.9) :

$$E(\hat{Y}_{ppz}) = \frac{1}{n}\sum_{i=1}^{N}(nz_i)\frac{y_i}{z_i} = \sum_{i=1}^{n} y_i = Y$$

أي أن $\hat{Y}_{ppz}$ غير منحاز. وأيضًا،

$$V(\hat{Y}_{ppz}) = \frac{1}{n^2}\left[\sum_{i=1}^{N}\left(\frac{y_i}{z_i}\right)^2 V(t_i) + 2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{y_i}{z_i}\frac{y_j}{z_j}\text{Cov}(t_i t_j)\right] \quad (9A.12)$$

$$= \frac{1}{n}\left[\sum_{i=1}^{N}\left(\frac{y_i}{z_i}\right)^2 z_i(1-z_i) - 2\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{y_i}{z_i}\frac{y_j}{z_j}z_i z_j\right] \quad (9A.13)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{N}\frac{y_i^2}{z_i} - Y^2\right) = \frac{1}{n}\sum_{i=1}^{N} z_i\left(\frac{y_i}{z_i} - Y\right)^2 \quad (9A.11)$$

باعتبار أن $\sum z_i = 1$ ، وهو المطلوب.

وبأخذ $z_i = M_i/M_0$ في النظرية (١٩ـ١٠) نجد النتائج الموافقة لمعاينة باحتمالات متناسبة مع الحجم.

ويمكن إعطاء عبارة بديلة لِـ $V(\hat{Y}_{ppz})$ فمن (9A.13) نجد،

$$nV(\hat{Y}_{ppz}) = \sum_{i=1}^{N}\frac{y_i^2(1-z_i)}{z_i} - 2\sum_{i=1}^{N}\sum_{j>i}^{N} y_i y_j \quad (9A.14)$$

وبما أن $1-z_i$ يساوي مجموع كافة المقادير z الأخرى في المجتمع، فيتضمن مُعامل y_i^2/z_i في (9A.14) حدًّا z_j لأي $j \neq i$. وبصورة مماثلة، يتضمن معامل y_j^2/z_j حدًّا z_i . وبالتالي،

$$V(\hat{Y}_{ppz}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N}\left(\frac{y_i^2 z_j}{z_i} + \frac{y_j^2 z_i}{z_j} - 2y_i y_j\right)$$

$$= \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N} z_i z_j\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 \quad (9A.15)$$

نظرية (١٩ ـ ٢).

إذا سحبنا، مع الإعادة، عينة تتضمن n وحدة باحتمالات تتناسب مع z_i ، فيكون،

$$v(\hat{Y}_{ppz}) = \sum_{i=1}^{n} \left(\frac{y_i}{z_i} - \hat{Y}_{ppz}\right)^2 \Big/ n(n-1) \qquad (9A.16)$$

تقدير عينة غير منحاز لِـ $V(\hat{Y}_{ppz})$، وذلك من أجل أي $n>1$.

برهان: من المطابقة الجبرية المعتادة نجد،

$$\sum_{i=1}^{n} \left(\frac{y_i}{z_i} - \hat{Y}_{ppz}\right)^2 = \sum_{i=1}^{n} \left(\frac{y_i}{z_i} - Y\right)^2 - n(\hat{Y}_{ppz} - Y)^2 \qquad (9A.17)$$

وبالتالي، من (9A.16) ، فإن

$$E[n(n-1)v(\hat{Y}_{ppz})] = E \sum_{i=1}^{n} \left(\frac{y_i}{z_i} - Y\right)^2 - nV(\hat{Y}_{ppz}) \qquad (9A.18)$$

ومن تعريف $V(\hat{Y}_{ppz})$ وبإدخال المتغيرات t_i ، مستخدمين (9A.11) في النظرية (١٩ـ١)، نجد

$$\begin{aligned} n(n-1)E[v(\hat{Y}_{ppz})] &= E \sum_{i=1}^{N} t_i\left(\frac{y_i}{z_i} - Y\right)^2 - nV(\hat{Y}_{ppz}) \\ &= n \sum_{i=1}^{n} z_i\left(\frac{y_i}{z_i} - Y\right)^2 - nV(\hat{Y}_{ppz}) \\ &= n(n-1)\,V(\hat{Y}_{ppz}) \end{aligned} \qquad (9A.19)$$

وهو المطلوب.

وبما أن $\hat{Y}_{ppz}$ هو متوسط n من القيم y_i/z_i فللعلاقة (9A.16) شكل بسيط جدًّا. وعندما يكون الاختيار متناسبًا حصرًا مع الحجم (أي أن $z_i=M_i/M_0$) فيمكن التعبير عن النظريتين (١٩ ـ ١) و(١٩ ـ ٢) كما يلي.

نظرية (١٩ ـ ٣)

إذا سحبنا، مع الإعادة، عيّنة تتضمن n وحدة باحتمالات تتناسب مع الحجم $z_i=M_i/M_0$ ، فعندئذ يكون

$$\hat{Y}_{pps} = \frac{M_0}{n} \sum_{i=1}^{n} \left(\frac{y_i}{M_i}\right) = \frac{M_0}{n} \sum_{i=1}^{n} (\bar{y}_i) = M_0\bar{\bar{y}} \qquad (9A.20)$$

حيث $\bar{y}$ المتوسط غير المرجّح لمتوسطات الوحدات، تقديرًا غير منحاز لـ Y بتباين

$$V(\hat{Y}_{pps}) = \frac{M_0}{n} \sum_{i=1}^{N} M_i (\bar{y}_i - \bar{\bar{Y}})^2 \qquad (9A.21)$$

وهذه النتائج تتبع من النظرية (١٩ ـ ١)، باعتبار أن $\bar{y}_i = y_i/M_i$ و $\bar{\bar{Y}} = Y/M_0.$

نظرية (١٩ ـ ٤)

تحت شروط النظرية (١٩ ـ ٣) يكون

$$v(\hat{Y}_{pps}) = M_0^2 \sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2 / n(n-1) \qquad (9A.22)$$

تقدير عيّنة غير منحاز لـ $V(\hat{Y}_{pps})$.

والنتيجة تتبع بتعويض $z_i = M_i/M_0$ ، في (9A.16) ، باعتبار أن $\bar{y}_i = y_i/N_i$ و $\hat{Y}_{ppz} = M_0 \bar{\bar{y}}.$

(١٩ ـ ٤) القياس الأمثل للحجم

في الحالات التي يكون فيها قياس الحجم M'_1 تقديرًا لكبر الوحدة، يبرز سؤال مهم هو: ما هو قياس الحجم الذي يجعل تباين $\hat{Y}_{ppz}$ أصغر ما يمكن؟

الآن،

$$V(\hat{Y}_{ppz}) = \frac{1}{n} \sum^{N} z_i \left(\frac{y_i}{z_i} - Y\right)^2 = \frac{1}{n} \left(\sum^{N} \frac{y_i^2}{z_i} - Y^2\right)$$

وتصبح هذه العبارة صفرًا إذا كان $z_i \alpha y_i$ وإذا كانت المقادير y_i جميعها موجبة، فهذه المجموعة من القيم z_i تشكل مجموعة مقبولة من الاحتمالات. وبالتالي فإن أفضل قياسات للحجم هي أعداد متناسبة مع مجاميع المفردات y_i في الوحدات المختلفة.

وهذه النتيجة ليست للتطبيق العملي المباشر؛ فإذا كانت المقادير y_i معروفة سلفًا في كامل المجتمع فسوف لا تكون العيّنة ضرورية. وتقترح النتيجة أنه إذا كانت المقادير y_i مستقرة نسبيًا مع الزمن، فقد تشكل أحدث قيم سابقة متوافرة لها أفضل

قياسات للحجم من أجل هذه المفردة. وعند اختيار عيّنة، في التطبيق العملي، يجب طبعًا ألا يختلف قياس الحجم المستخدم من مفردة إلى أخرى. وإذا كان هناك مجال للاختيار بين قياسين مختلفين للحجم، فمن المرجّح أن يكون القياس الأفضل هو القياس الأكثر قربًا إلى وضع التناسب مع مجاميع المفردات الرئيسة في الوحدات.

(١٩ ـ ٥) الدقة النسبية لثلاث طرق

نقارن في هذه الفقرة دقة الطرق الثلاث السابقة لتقدير مجموع المجتمع مع وحدات عنقودية ذات حجوم غير متساوية (مفترضين أن M_i معروفة إذا استدعت الطريقة ذلك).

١ ـ اختيار: احتمالات متساوية. تقدير؛ $\hat{Y}_u$.

٢ ـ اختيار: احتمالات متساوية. تقدير؛ $\hat{Y}_R$.

٣ ـ اختيار: احتمال متناسب مع الحجم. تقدير؛ $\hat{Y}_{pps}$.

ولا توجد قاعدة بسيطة لتحديد الطريقة الأكثر دقة بين الطرق الثلاث. وتعتمد المسألة على العلاقة بين $\bar{y}_i$ و M_i وعلى تباين $\bar{y}_i$ كدالّة في M_i. والحالة الملائمة لتقديري النسبة والـ ا م ح (أ م ح ترمز لاحتمال متناسب مع الحجم) هي تلك التي لا يكون $\bar{y}_i$ فيها على صلة بـ M_i. والحالة الملائمة لـ $\hat{Y}_u$ هي تلك التي لا يكون مجموع الوحدة y_i فيها على صلة بـ M_i.

ويمكن الحصول على بعض الدليل بالتعبير عن تباينات التقديرات الثلاثة في شكل قابل للمقارنة. ونفترض أن $(N\text{-}1) \doteq N$ كما نكتب $E(y_i-\bar{Y})^2=\sum^N (y_i-\bar{Y})^2/N$ ونفترض أيضًا أنه يمكن إهمال الانحياز في $\hat{Y}_R$.

وفي حالة $\hat{Y}_u$ نجد، من (9A.2) أن،

$$nV(\hat{Y}_u)=N^2(1-f)E(y_i-\bar{Y})^2=(1-f)E(Ny_i-Y)^2 \qquad (9A.23)$$

أما في حالة $\hat{Y}_R$ فنجد، من (9A.4) ، أن،

$$nV(\hat{Y}_R)=N^2(1-f)EM_i^2(\bar{y}_i-\bar{\bar{Y}})^2=(1-f)E\left(\frac{M_i}{\bar{M}}\right)^2(M_0\bar{y}_i-Y)^2 \qquad (9A.24)$$

حيث $\bar{M} = \sum M_i/N = M_0/N$.

ومن (9A.21) نجد، من أجل $\hat{Y}_{pps}$ أن،

$$nV(\hat{Y}_{pps}) = NM_0EM_i(\bar{y}_i - \bar{\bar{Y}})^2 = M_0{}^2E\left(\frac{M_i}{\bar{M}}\right)(\bar{y}_i - \bar{\bar{Y}})^2$$

$$= E\left(\frac{M_i}{\bar{M}}\right)(M_0\bar{y}_i - Y)^2 \qquad (9A.25)$$

ومن (9A.23), (9A.24) و (9A.25) نرى أن $V(\hat{Y}_u)$ يعتمـد على دقة المقادير $Ny_i = NM_i\bar{y}_i$ كتقديرات لِـ Y ، بينما يعتمد $V(\hat{Y}_R)$ و $V(\hat{Y}_{pps})$ على دقة المقادير $M_0\bar{y}_i = M_0y_i/M_i$ كتقديرات لِـ Y ، وإذا لم يكن $\bar{y}_i$ على صلة بِـ M_i فنتوقع أن يكون $M_0\bar{y}_i$ أكثر دقة من الـ $NM_i\bar{y}_i$، ونتوقع العكس إذا لم يكن y_i على صلة بِـ M_i .

وبـالنسبـة لِـ $\hat{Y}_R$ و $\hat{Y}_{pps}$ نلاحظ من (9A.24) و (9A.25) أن $V(\hat{Y}_R)$ يعـطي للوحـدات الكبـيرة ترجيحـة أكـبر نسبيًا مما يعـطيه $V(\hat{Y}_{pps})$. ونـلاحظ أيضًـا أن $\hat{Y}_u$ و $\hat{Y}_R$ يستفيدان من حد الـ ت م م الذي يمكن أن يصبح كبيرًا في طبقات صغيرة (مثلاً عندما $n_h = 2, N_h = 10$).وقد حَدت هذه النقطة بتطوير الاختيار باحتمالات غير متساوية بدون إعادة. وتصحّ العلاقة (9A.24) الخاصة بِـ $\hat{Y}_R$ ، بالطبع، في العيّنات الكبيرة فقط.

وقد جرت مقارنات إضافية بين الطرق الثلاث وفق نموذج مجتمع لا نهائي من قِبـل Cochran الطبعة الثانية، Des Raj (1954, 1958), Yates (1960), Zarcovic (1960) و Foreman و Brewer (1971) . ويفـترض معـظم الكتّاب أن المجتمع المنتهي عيّنة عشوائية من مجتمع فوقي لا نهائي يكون فيه،

$$y_i = \alpha + \beta M_i + e_i; \qquad E(e_i|M_i) = 0 \qquad (9A.26)$$

مما يؤمل أنه يشكل تقريبًا للعلاقة التي تصحّ في العديد من المسوح الإحصائية. ويجب أيضًا وضع بعض الفروض حول تباين e_i في عناقيد ذات حجم معطى.

ومن (9A.12) في الفقرة (٩ ـ ٤)، نجد بعد القسمة على $N \doteq (N\text{-}1)$ ،

$$V(e_i) = V(y_i|M_i) \doteq M_i S^2[M_i\rho + (1-\rho)]$$

وكتقـريب يقـترح هذا كون $V(e_i) = cM_i^g$ حيث $1 < g < 2$ في معـظم التـطبيقات. ومن النموذج (9A.26) نجد،

$$\bar{Y} = \alpha + \beta\bar{M} + \bar{e}_N: \qquad \bar{\bar{Y}} = \frac{\alpha}{\bar{M}} + \beta + \frac{\bar{e}_N}{\bar{M}} \tag{9A.27}$$

ونفترض هنا أنه يمكن إهمال $\bar{e}_N$ ؛ وهذا يؤدي إلى تجاهل الـ ت م م. ونستنتج أن،

$$\frac{nV(\hat{Y}_u)}{N^2} = E(y_i - \bar{Y})^2 = E[\beta(M_i - \bar{M}) + e_i]^2$$

$$= \beta^2 V(M_i) + cE(M_i^g) \tag{9A.28}$$

$$\frac{nV(\hat{Y}_R)}{N^2} \doteq EM_i^2(\bar{y}_i - \bar{\bar{Y}})^2 = E(y_i - M_i\bar{\bar{Y}})^2$$

$$= E[\alpha(1 - M_i/\bar{M}) + e_i]^2 = \frac{\alpha^2 V(M_i)}{\bar{M}^2} + cE(M_i^g) \tag{9A.29}$$

$$\frac{nV(\hat{Y}_{pps})}{N^2} \doteq \bar{M}EM_i(\bar{y}_i - \bar{\bar{Y}})^2 = \bar{M}EM_i\left[\alpha\left(\frac{1}{M_i} - \frac{1}{\bar{M}}\right) + \frac{e_i}{M_i}\right]^2$$

$$= \alpha^2 E\left[\frac{(M_i - \bar{M})^2}{M_i\bar{M}}\right] + v\bar{M}E(M_i^{g-1}) \doteq \frac{\alpha^2 V(M_i)}{\bar{M}^2} + c\bar{M}E(M_i^{g-1}) \tag{9A.30}$$

وبالتالي، تكون التباينات القابلة للمقارنة، بصورة تقريبية، تحت هذا النموذج هي:

$$\frac{nV_u}{N^2} = \beta^2 V(M_i) + cE(M_i^g) \tag{9A.31}$$

$$\frac{nV_R}{N^2} \doteq \alpha^2 \frac{V(M_i)}{\bar{M}^2} + cE(M_i^g) \tag{9A.32}$$

$$\frac{nV_{pps}}{N^2} \doteq \frac{\alpha^2 V(M_i)}{\bar{M}^2} + c\bar{M}E(M_i^{g-1}) \tag{9A.33}$$

لنعتبر أولاً $\alpha = 0$، الحالة التي لا يكون فيها $\bar{y}_i$ على صلة بِـ M_i. فمن الواضح أن،

$$V_R < V_u$$

وإذا كان β كبيراً (ويصبح في هذه الحالة $\bar{\bar{Y}}$)، فقد يكون تفوّق التقدير النسبة كبيراً وفي حالة $\alpha = 0$ يكون،

$V_{pps} < V_u$ من أجل $g \geq 1$

ويصحّ هذا بسبب أنه إذا كان $g > 1$ فالتغاير بين M_i^{g-1} و M_i موجب إذا كان $g \geq 1$، وهو صفر إذا كان $g = 1$، وهكذا نجد في حالة $g \geq 1$، أن

$$E(M_i^g) = E[(M_i)(M_i^{g-1})] \geq \bar{M}E(M_i^{g-1})$$

وإذا كان $\beta = 0$ و $\alpha \neq 0$ بحيث لا يكون مجموع الوحدة $\bar{y}_i$ على صلة بـ M_i فإن $\hat{Y}_u$ يهزم دائمًا $\hat{Y}_R$ و $\hat{Y}_{pps}$، ربما باستثناء الحالة غير محتملة الوقوع $g = 2$ و $\beta = 0$.

وعندما لا ينعدم α ولا β يعتمد الإنجاز النسبي لـ $\hat{Y}_u$، $\hat{Y}_R$ و $\hat{Y}_{pps}$، على الحجوم النسبية لـ $|\alpha|$ و $|\beta|$. وعلى سبيل المثال، يهزم $\hat{Y}_R$ التقدير $\hat{Y}_u$ إذا كان $\beta^2 > \alpha^2/\bar{M}^2$، كما لاحظ Foreman و Brewer (1971).

وفيما يتعلق بـ V_R في مقابل V_{pps} نجد أن معامل α^2 هو نفسه تقريبًا في V_R و V_{pps}. ومن الحدود التي تتضمن c يُتوقع أن نجد في معظم التطبيقات التي يكون فيها $\beta \neq 0$ أن $V_{pps} < V_R$ في حالة $g > 1$ و $V_{pps} > V_R$ في حالة $g < 1$.

وتتفق نتائج هذا النموذج مع النتائج المقترحة منذ بداية هذه الفقرة. فإذا كان $\bar{y}_i$ لا يُظهر اتجاهًا معتدلاً مع تزايد M_i فإن طريقة النسبة باحتمالات متساوية وطريقة الـ pps تتفوقان على التقدير غير المنحاز باحتمالات متساوية، وقد يكونان أكثر دقة بكثير. ويتفوق $\hat{Y}_u$ إذا لم يكن مجموع الوحدة y_i على صلة بـ M_i. ويوجد القليل من الاختيار بين التقدير النسبة وتقدير الـ ا م ح. وبما أننا نتوقع وقوع g بين 1 و 2 فيكون تقدير الـ ا م ح، عادة، أكثر دقة، ونتائجه ليست مقصورة على العيّنات الكبيرة.

ويساعد حدّ الـ ت م م التقديرين $\hat{Y}_u$ و $\hat{Y}_R$ عندما يكون هذا الحد غير مهمل. وبانتظار عمل لاحق (فقرة ١٩ ـ ١٢)، تقترح الدلالات المتوافرة هنا أن أفضل طرق الـ ا م ح بدون إعادة تستفيد من حوالي نفس حجم حد الـ ت م م استفادة المعاينة مع احتمالات متساوية.

(١٩ ـ ٦) المعاينة باحتمالات غير متساوية دون إعادة

أُنتج الكثير من هذا العمل لخدمة مسوح إحصائية واسعة قُسمت فيها الوحدات العنقودية أولاً، ووفق مبدأ آخر (مثلاً، الموقع الجغرافي) إلى عدد كبير من الطبقات الصغيرة نسبيًا، ثم سُحب عدد صغير فقط من الوحدات العنقودية من كل طبقة. والحالة $n_h = 2$، التي تقدم درجة حرية واحدة من كل طبقة لتقدير أخطاء المعاينة، هي حالة ذات أهمية خاصة. لنفرض أننا سحبنا وحدتين من طبقة، وأن الوحدة الأولى مسحوبة باحتمالات z_j، تتناسب مع قياس ما للحجم. ولتكن الوحدة i هي الوحدة المختارة. فإذا اتبعنا الطريقة الأكثر بداهة، فإننا نختار في السحب الثاني إحدى الوحدات الباقية بعد تخصيص احتمالات هي $z_j/(1-z_i)$. وبالتالي يكون الاحتمال الكلي لاختيار الوحدة i في أي من السحبين الأول أو الثاني هو،

$$\pi_i = z_i + \sum_{j \neq i}^{N} \frac{z_j z_i}{(1-z_j)} = z_i\left(1 + \sum_{j \neq i}^{N} \frac{z_j}{1-z_j}\right) \qquad (9A.34)$$

$$= z_i\left(1 + A - \frac{z_i}{1-z_i}\right) \qquad (9A.35)$$

حيث $A = \sum z_j/(1-z_j)$ مأخوذًا فوق الوحدات الـ N جميعها.

لنفرض أن $\pi_i = 2z_i$ فالاحتمالات النسبية لاختيار الوحدات الباقية متناسبة مع قياس للحجم هو z_i، ومقدّر العيّنة الذي قدّمه Horvitz و Thompson (1952)، وهو

$$\hat{Y}_{HT} = \sum_{i}^{2} \frac{y_i}{\pi_i} = \frac{1}{2}\sum_{i}^{2} \frac{y_i}{z_i}$$

سيكون له عندئذ تباين يساوي الصفر إذا كانت المقادير z_i متناسبة مع المقادير y_i، باعتبار أن $z_i = y_i/Y$، وستعطي كل وحدة اختيرت التقدير الصحيح $y_i/z_i = Y$. إلا أن المقادير $z'_i = \pi_i/2$، في (9A.35)، هي دائمًا أقرب إلى التساوي من المقادير z_i الأصلية بسبب من العامل الثاني في (9A.35). وفي المثال الذي أعطاه Yates و Grundy (1953)، مع $n=2, N=4$ و z_i تساوي 0.3, 0.2, 0.1 و 0.4، وُجد أن z'_i تساوي 0.3042, 0.2206, 0.1173 و 0.3579.

وقد استُخدمت ثلاث طرق لمكافحة هذا الالتواء فيما نعتزم أن تكونه احتمالات الاختيار. واحدة نحتفظ فيها بالطريقة الطبيعية السابقة لاختيار العيّنة، ولكننا نقوم بتغيرات مناسبة في طريقة تقدير مجموع مجتمع. وواحدة نستخدم فيها طريقة مختلفة لاختيار العيّنـة، وتحفظ هذه الـطريقة π_i مساوية لِـ nz_i تحت قيود معيّنة نضعها على قيم nz_i. وواحـدة هي قبول بعض الالتواء في الاحتمالات إذا كان له ميزات مجزية (مثلاً، في مجال البساطة أو الشمولية). وسنعطي في الفقرة (١٩ ـ ٨) وما بعدها أمثلة على الطرق الثلاث.

ونقدّم أولاً التقدير العام الأكثر شهرة لمجموع مجتمع في حالة معاينة باحتمالات غير متساوية وبدون إعادة.

(١٩ ـ ٧) مقدّر هيرفتز ـ تومبسون (Horvitz-Thompson)

اختيرت عيّنة من n وحدة، بدون إعادة، وفق طريقة ما. ليكن،

π_i = احتمال أن تكون الوحدة i ضمن العيّنة.

π_{ij} = احتمال أن تكون الوحدتان i و j كلاهما ضمن العيّنة.

فتصحّ عندئذ العلاقات التالية :

$$\sum_{i}^{N} \pi_i = n, \qquad \sum_{j \neq i}^{N} \pi_{ij} = (n-1)\pi_i, \qquad \sum_{i}^{N} \sum_{j>i} \pi_{ij} = \tfrac{1}{2}n(n-1) \qquad (9A.36)$$

ولإرسـاء العـلاقـة الثـانية، لنرمز بِـ $P(s)$ لاحتمال أن تتضمن عيّنة n من الوحدات المحـدّدة. فعنـدئـذ يكـون $\pi_{ij} = \Sigma P(s)$ فوق جميع العيّنـات التي تتضمن الوحدتين الـ i والـ j ، و $\pi_i = \Sigma P(s)$ فوق جميع العيّنات التي تتضمن الوحدة i . وعندما نأخذ $\Sigma\pi_{ij}$ حيث $j \neq i$ ، فإن كل $P(s)$ لعـيّنـة تحوي الـوحـدة i يتـكـرر تعـدادهـا في المجموع $(n-1)$ مرة، ذلك لأن العيّنة تتضمن $(n-1)$ من القيم الممكنة لِـ j . وهذا يثبت العلاقة الثانية. أما العلاقة الثالثة فتتبع من الثانية.

ومقدّر Horvitz-Thompson (1952) لمجموع المجتمع هو،

$$\hat{Y}_{HT} = \sum_{i}^{n} \frac{y_i}{\pi_i} \qquad (9A.37)$$

حيث y_i القياس من الوحدة i .

نظرية (١٩ ـ ٥)

إذا كان $\pi_i > 0$ ، $(i=1,2,\ldots,N)$ ، فعندئذ يكون ،

$$\hat{Y}_{HT} = \sum_{i}^{n} \frac{y_i}{\pi_i}$$

تقديرًا غير منحاز لِـ Y ، بتباين :

$$V(\hat{Y}_{HT}) = \sum_{i=1}^{N} \frac{(1-\pi_i)}{\pi_i} y_i^2 + 2 \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} y_i y_j \qquad (9A.38)$$

حيث π_{ij} احتمال أن تكون الوحدتان i و j ضمن العيّنة .

برهان :

ليكن $t_i (i=1,2,\ldots,N)$ متغيرًا عشوائيًا يأخذ القيمة 1 إذا سُحبت الوحدة i ، ويكون صفرًا فيما عدا ذلك. فعندئذ يتبع t_i التوزيع الثنائي لعيّنة حجمها 1 ، باحتمال يساوي π_i ، وهكذا يكون ،

$$E(t_i) = \pi_i, \qquad V(t_i) = \pi_i(1-\pi_i) \qquad (9A.39)$$

ونحتاج أيضًا إلى قيمة $\text{Cov}\,(t_i t_j)$. وبما أن $t_i t_j$ لا يساوي 1 إلا إذا ظهرت الوحدتان كلتاهما في العيّنة ، فلدينا

$$\text{Cov}\,(t_i t_j) = E(t_i t_j) - E(t_i)E(t_j) = \pi_{ij} - \pi_i \pi_j \qquad (9A.40)$$

وبالتالي نجد ، معتبرين الـ y_i كمقادير مثبتة والـ t_i متغيرات عشوائية ،

$$E(\hat{Y}_{HT}) = E\left(\sum_{i=1}^{N} \frac{t_i y_i}{\pi_i}\right) = \sum_{i=1}^{N} y_i = Y$$

$$V(\hat{Y}_{HT}) = \sum_{i}^{N} \left(\frac{y_i}{\pi_i}\right)^2 V(t_i) + 2 \sum_{i}^{N} \sum_{j>i}^{N} \frac{y_i}{\pi_j} \frac{y_j}{\pi_j} \text{Cov}\,(t_i t_j)$$

$$= \sum_{i}^{N} \frac{(1-\pi_i)}{\pi_i} y_i^2 + 2 \sum_{i}^{N} \sum_{j>i}^{N} \frac{(\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} y_i y_j \qquad (9A.41)$$

وهو المطلوب .

ويمكن التعبير عن هذا التباين بشكل آخر مستخدمين العلاقتين الأولى والثانية في (9A.36) . وهذا يعطي ،

$$\sum_{j\neq i}(\pi_{ij}-\pi_i\pi_j)=(n-1)\pi_i-\pi_i(n-\pi_i)=-\pi_i(1-\pi_i)$$

وبالتعويض عن $(1-\pi_i)$ في الحد الأول من (9A.41) نجد ،

$$\sum_i^N\frac{(1-\pi_i)}{\pi_i}y_i^2=\sum_i^N\sum_{j\neq i}^N(\pi_i\pi_j-\pi_{ij})\left(\frac{y_i}{\pi_i}\right)^2=\sum_i^N\sum_{j>i}^N(\pi_i\pi_j-\pi_{ij})\left[\left(\frac{y_i}{\pi_i}\right)^2+\left(\frac{y_j}{\pi_j}\right)^2\right]$$

ومنه ،

$$V(\hat{Y}_{HT})=\sum_i^N\sum_{j>i}^N(\pi_i\pi_j-\pi_{ij})\left[\left(\frac{y_i}{\pi_i}\right)^2+\left(\frac{y_j}{\pi_j}\right)^2-2\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}\right]$$

$$=\sum_i^N\sum_{j>i}^N(\pi_i\pi_j-\pi_{ij})\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \qquad (9A.42)$$

نتيجة

من (9.41) ، مستخدمين طريقة الـ t_i ، نجد تقدير عيّنة غير منحاز لِـ $V(\hat{Y}_{HT})$ هو

$$v_1(\hat{Y}_{HT})=\sum_i^n\frac{(1-\pi_i)}{\pi_i^2}y_i^2+2\sum_i^n\sum_{j>i}^n\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j\pi_{ij}}y_iy_j \qquad (9A.43)$$

شريطة ألا ينعدم أي من الـ π_{ij} في المجتمع .

وقد أعطى Yates و Grundy (1953) و Sen (1953) تقدير عيّنة مختلف . ومن (9A.42) نجد هذا التقدير على الشكل ،

$$v_2(\hat{Y}_{HT})=\sum_i^n\sum_{j>i}^n\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \qquad (9A.44)$$

مع القيود نفسها على المقادير π_{ij} .

وبما أن الحدود $(\pi_i\pi_j-\pi_{ij})$ تختلف في الغالب اختلافًا كبيرًا بعضها عن بعض ، وتكون أحيانًا سالبة ، فإن المقدارين v_1 و v_2 ينزعان إلى عدم الاستقرار. ويمكن أن

يفترض كلا التقديرين قيمًا سالبة في بعض طرق اختيار العيّنة. وقد قارن Rao و Singh (1973) معامل اختلاف v_1 و v_2 في عيّنات حجمها $n=2$ من 34 مجتمعًا فعليًا صغيرًا عثروا عليها في كتب وأبحاث تتعلق بمسوح عيّنات إحصائية، وقد استخدما طريقة Brewer (فقرة ١٩ ـ ٨)، في اختيار العيّنة، التي يكون فيها $\pi_i = 2z_i$ كما هو مرغوب. وكان التقدير v_2 أكثر استقرارًا بكثير بالإضافة إلى كونه غير سالب على الدوام في هذه الطريقة، بينما تكرّر اتخاذ v_1 لقيم سالبة.

ونعود الآن إلى طرق اختيار عيّنة. إذ يذكر ثَبْت المراجع الذي أعدّه Brewer و Hanif (1969) ما يزيد على 30 طريقة، سنستعرض منها أربع طرق. وتصبح معظم الطرق أكثر تعقيدًا باطراد عندما تتجاوز n القيمة 2 ، والقليل منها فقط قابل للتعميم بسهولة. وسيكرّس العَرض هنا معظم الإنتباه إلى الحالة $n=2$ ، نظرًا للبساطة ولأنها الحالة السائدة في عيّنات، على مستوى قومي، تستخدم العديد من الطبقات الصغيرة مع $n_h=2$.

(١٩ ـ ٨) طريقة بروير Brewer

في حالة $n=2$ تحافظ طريقة اختيار العيّنة هذه على $\pi_i = 2z_i$ وتستخدم مقدّر Horvitz -Thompson ،

$$\hat{Y}_{HT} = \frac{y_i}{\pi_i} + $$

وباستخدام أساليب مختلفة، أعطت طرق أنتجها Brewer (1963) ، Rao (1965) و Durbin (1967) القيم نفسها لِـ π_i و π_{ij} . ونفترض كل z_i أصغر من نصف.

ويسحب Brewer الوحدة الأولى بما يسميه الاحتمالات المحسّنة المتناسبة مع $z_i(1-z_i)/(1-2z_i)$ والوحدة الثانية باحتمالات $z_i/(1-z_j)$ حيث j الوحدة المسحوبة أولاً. والمقام اللازم لجعل المقادير $z_i(1-z_i)/(1-2z_i)$ احتمالات فعلية هو مجموعها

$$D = \sum_{i=1}^{N} \frac{z_i(1-z_i)}{1-2z_i} = \frac{1}{2}\sum_{i=1}^{N} \frac{z_i(2-2z_i)}{1-2z_i} = \frac{1}{2}\left(1+\sum_{i}^{N} \frac{z_i}{1-2z_i}\right) \qquad (9A.45)$$

ومع n=2 يكون احتمال سحب الوحدة i هو مجموع احتمالي سحبها أولاً وسحبها ثانيًا. وهكذا نجد

$$\pi_i = \frac{z_i(1-z_i)}{D(1-2z_i)} + \frac{1}{D}\sum_{j\neq i}^{N} \frac{z_j(1-z_j)}{(1-2z_j)}\frac{z_i}{(1-z_j)}$$
$$= \frac{z_i}{D}\left[\frac{(1-2z_i)+z_i}{1-2z_i} + \sum_{j\neq i}^{N}\frac{z_j}{1-2z_j}\right] = \frac{z_i}{D}\left(1+\sum_{j=1}^{N}\frac{z_j}{1-2z_j}\right) = 2z_i \quad (9A.46)$$

متذكرين (9A.45) من أجل D . وبصورة مماثلة ،

$$\pi_{ij} = \frac{z_i z_j}{D}\left(\frac{1}{1-2z_i} + \frac{1}{1-2z_j}\right) = \frac{2z_i z_j}{D}\frac{(1-z_i-z_j)}{(1-2z_i)(1-2z_j)} \quad (9A.47)$$

وبما أن هذه الطريقة تستخدم التقدير HT ، فالنظرية (٩ـ٥) ونتيجتها تقدمان العلاقات الخاصة بتباين $\hat{Y}_B$ وبتقدير لهذا التباين. ولهذه الطريقة خاصتان مرغوبتان. فقد بيّن Brewer (1963) أن تباينه أقل دائمًا من تباين التقدير $\hat{Y}_{PPS}$ في معاينة مع الإعادة. وثانيًا تبيّن بعض العمليات الجبرية (1965, Rao) أن $0 < (\pi_i\pi_j - \pi_{ij})$ من أجل جميع القيم $i\neq j$ ، وهكذا يكون التقدير v_2 للتباين، وهو تقدير Yates-Grundy ، موجبًا دومًا.

ووفق أسلوب Durbin (1967) نسحب الوحدة الأولى (i) باحتمال z_i ، وإذا سحبنا الوحدة i أولاً نجعل احتمال سحب الوحدة j ثانيًا متناسبًا مع

$$z_j\left[\frac{1}{(1-2z_i)} + \frac{1}{(1-2z_j)}\right] \quad (9A.48)$$

وفي هذه الحالة يكون مقام النسب،

$$\sum_{j\neq i}^{N} z_j\left[\frac{1}{(1-2z_i)} + \frac{1}{(1-2z_j)}\right] = \frac{(1-z_i)}{1-2z_i} + \sum_{j\neq i}^{N}\frac{z_j}{(1-2z_j)} = 1 + \sum_{j=1}^{N}\frac{z_j}{1-2z_j} \quad (9A.49)$$

وهكذا يكون المقام مساويًا لِـ 2D في علاقة Brewer ,(9A.45) . واحتمال سحب الوحدة i أولاً والوحدة j ثانيًا هو إذن،

$$P(i)P(j|i) = \frac{z_i z_j}{2D}\left[\frac{1}{(1-2z_i)} + \frac{1}{(1-2z_j)}\right] \quad (9A.50)$$

ومن التناظر يتساوى هذا مع $P(j)P(i|j)$ أي أن π_{ij} في حالة Durbin هو π_{ij} ذاته في حالة Brewer (9A.47) .

وقد عمّم Sampford (1967) هذه الطريقة إلى عيّنات حجمها n ، شريطة أن يكون $nz_i < 1$ في جميع وحدات المجتمع. وبهذه الطريقة في اختيار العيّنة يكون احتمال أن تتألف العينة، مثلاً، من الوحدات $1,2,...,n$ ، تعميمًا طبيعيًا لـ (9A.47) ، أي متناسبًا مع :

$$\left(1-\sum_{i=1}^{n} z_i\right) \prod_{i=1}^{n} z_i \Big/ \prod_{i=1}^{n} (1-nz_i) \qquad (9A.51)$$

وفي هذه الطريقة يمكن البرهان على أن $\pi_i = nz_i$. وهناك علاقة معطاة من أجل π_{ij} مع نصيحة باستخدام الحاسب الإلكتروني في حسابها. ويُستخدم المقدّر HT لتقدير Y ، وبالتالي هناك علاقات متوافرة، تعطي تباينه وتقدير هذا التباين. وتقدير Yates-Grundy وهو التقدير v_2 المذكور في (9A.44) ، موجب دومًا. وقد اقترح Sampford عدة طرق عملية لسحب العيّنة بحيث تحقق (9A.51) . وإحداها هي سحب الوحدة الأولى باحتمالات z_i وجميع الوحدات التالية باحتمالات متناسبة مع $z_i/(1-nz_i)$ مع الإعادة. وسيكون هذا مقبولاً إذا حصلنا على عيّنة فيها n من الوحدات المتميزة. ونرفض محاولة عيّنة حالما نحصل على وحدة مرتين. ويمكن رؤية أن هذه الطريقة تقود إلى (9A.51) . وكمرشد لسرعة هذه الطريقة قُدّمت علاقة تعطي العدد المتوقع للمحاولات المطلوبة حتى نحصل على عيّنة.

وفي حالة $n=2$ ، تتمتع طريقة Durbin (1967) لسحب عيّنة، وخلافًا لطريقة Brewer ، بخاصة أن الاحتمال غير المشروط لسحب وحدة i هو z_i في كل من السحبين الأول والثاني. وفي المعاينة متعددة المراحل في مسوح تتكرر وفق فترات زمنية منتظمة، أشار Felligi (1963) إلى أنه من الضروري أو المستحسن حذف وحدات واستبدالها من وقت لآخر وفق نمطية منتظمة فيما يُسمّى خطة دورانية، باعتبار أن الاستمرار الطويل في سؤال الأشخاص أنفسهم أمر غير مرغوب. وقد أعطى طريقة لاختيار الوحدات المتتالية تمتلك أيضًا الخاصة التي تمتلكها طريقة Durbin . وطريقته، وهي قائمة على حسابات متكررة، مشابهة لطريقة *Durbin-Brewer* إلا أن π_{ij} مختلفة قليلاً.

(١٩ـ٩) طريقة مورثي (Murthy)

تستخدم هذه الطريقة أسلوب الاختيار المقترح أولاً (فقرة ١٩ ـ ٦)، وتُسحب الوحدات المتتالية باحتمالات $z_i, z_j/(1-z_i), z_k/(1-z_i-z_j)$، وهكذا. ويتبّع مقدّر Murthy (1957) خُطى عمل سابق قام به Des Raj (1956) ، الذي قدّم تقديرًا بارعًا غير منحاز قائمًا على الترتيب المحدّد الذي سُحبت بموجبه وحدات العيّنة الـ n . وقد بيّن Murthy أنه في مقابل أي تقدير مرتّب من هذا الفصل من التقديرات يمكن إقامة تقدير غير مرتّب، وغير منحاز أيضًا وله تباين أصغر.

ومقدِّره المقترح هو،

$$\hat{Y}_M = \frac{\sum_i^n P(s|i)y_i}{P(s)} \qquad (9A.52)$$

حيث،

$P(s/i)$ = الاحتمال الشرطي للحصول على مجموعة الوحدات المسحوبة، علمًا أن الوحدة الـ i قد سُحبت أولاً .

$P(s)$ = الاحتمال غير الشرطي للحصول على مجموعة الوحدات المسحوبة .

ونبرهن الآن أن التقدير $\hat{Y}_M$ غير منحاز فلأي وحدة i في المجتمع، $\sum P(S|i)=1$ حيث يمتد المجموع فوق جميع العيّنات التي سحبت فيها الوحدة i أولاً . لبرهان هذا في حالة $n=2$ حيث j الوحدة الأخرى في العيّنة، نكتب

$$P(s|i) = \frac{z_j}{1-z_i}; \qquad \sum P(s|i) = \frac{\sum_{j\neq i}^{N} z_j}{1-z_i} = 1$$

وفي حالة $n=3$ ، حيث j و k الوحدتان الثانية والثالثة، نجد،

$$\sum P(s|i) = \sum_{j\neq i}^{N} \sum_{k\neq i,j}^{N} z_j z_k/(1-z_i)(1-z_i-z_j) = \sum_{j\neq i}^{N} z_j/(1-z_i) = 1$$

وهكذا من أجل $n>3$. وبالتالي، عندما نجمع $\sum P(s)\hat{Y}_M$ فوق جميع العيّنات ذات الحجم n ، نجد أن معامل y_i في هذا المجموع يساوي الواحد، بحيث يكون

$$E(\hat{Y}_M)=\sum P(s)\,\hat{Y}_M=\sum_i^N y_i=Y \tag{9A.53}$$

وقد أعطى Murthy (1957) عبارات عامة لـ $v(\hat{Y}_M)$ و $V(\hat{Y}_M)$ ومهما يكن n .

وعندما يكون $n=2$ ، تتضمن العينة الوحدتين i و j،

$$P(s|i)=\frac{z_j}{1-z_i};\qquad P(s|j)=\frac{z_i}{1-z_j} \tag{9A.54}$$

$$P(s)=\pi_{ij}=z_iP(s|i)+z_jP(s|j)=z_iz_j(2-z_i-z_j)/(1-z_i)(1-z_j) \tag{9A.55}$$

ويكون التقدير إذن، مستخدمين (9A.52) ، (9A.53) و (9A.54) على الشكل،

$$\hat{Y}_M=\frac{1}{2-z_i-z_j}\left[(1-z_j)\frac{y_i}{z_i}+(1-z_i)\frac{y_j}{z_j}\right] \tag{9A.56}$$

وبالتعريف يكون $V(\hat{Y}_M)$ (في حالة $n=2$) :

$$\sum_s P(s)\,\hat{Y}_M{}^2-Y^2=\sum_i^N\sum_{j>i}^N\frac{z_iz_j(2-z_i-z_j)}{(1-z_i)(1-z_j)}\,\hat{Y}_M{}^2-Y^2 \tag{9A.57}$$

وبعد التعويض عن $\hat{Y}_M{}^2$ من (9A.56) وعن Y^2، وإعادة الترتيب، نجد،

$$V(\hat{Y}_M)=\sum_i^N\sum_{j>i}^N\frac{z_iz_j(1-z_i-z_j)}{2-z_i-z_j}\left(\frac{y_i}{z_i}-\frac{y_j}{z_j}\right)^2 \tag{9A.58}$$

وبالمقارنة مع $V(\hat{Y}_{ppz})$ لمعاينة مع الإعادة المعطى في العلاقة (9A.15) ، نجد بوضوح أن $V(\hat{Y}_M)<V(\hat{Y}_{ppz})$.

وبما أن المتوسط هو $\sum P(s)v(\hat{Y}_M)$ من أجل أي $v(\hat{Y}_M)$ مقترح، فيكون

$$v(\hat{Y}_M)=\frac{(1-z_i)(1-z_j)(1-z_i-z_j)}{(2-z_i-z_j)^2}\left(\frac{y_i}{z_i}-\frac{y_j}{z_j}\right)^2 \tag{9A.59}$$

تقدير عيّنة غير منحازة للتباين في حالة $n=2$. وهو موجب دومًا.

والعلاقة الموافقة لعيّنات حجمها n هي،

$$v(\hat{Y}_M)=\frac{1}{[P(s)]^2}\left\{\sum_i^n\sum_{j>i}^n[P(s)P(s|i,j)-P(s|i)P(s|j)]z_iz_j\left(\frac{y_i}{z_i}-\frac{y_j}{z_j}\right)^2\right\} \tag{9A.60}$$

حيث $P(s|ij)$ هو الاحتمال الشرطي للحصول على عيّنة علمًا بأن الوحدتين i و j قد اختيرتا (بأي ترتيب) في السحبين الأول والثاني ، وقد أعطى Bayless (1968) برنامج حاسب لحساب الـ $P(s|i)$ و $P(s|ij)$.

(١٩ ـ ١٠) طرق لها صلة بالمعاينة النمطية

اقترح Madow (1949) طريقة ، يلاحظ Murthy (1967) أنها استُخدمت في مسوح إحصائية في الهند ، وهي امتداد للمعاينة النمطية . وفوائدها هي أن سحب العيّنة سهل مهما يكن n ، وهي تحافظ على π_i مساوية لِـ nz_i ، كما تقدم تقديرًا غير منحاز لِـ Y . وفضلاً عن ذلك ، فقد تتوافر للمعاين أحيانًا معلومات مسبقة كافية لترتيب الوحدات ، بصورة تقريبية على الأقل، ترتيبًا تؤدي معه المعاينة النمطية أداءً حسنًا . والمطعن في الطريقة النمطية هو كالمعتاد عدم توافر تقدير غير منحاز $v(\hat{Y}_{sys})$ للتباين $V(\hat{Y}_{sys})$.

ويمكن سحب عيّنة حجمها n مستخدمين إما الـ z_i أو قياسات الحجم M_i' التي اشتُقّت منها المقادير $z_i = M_i'/\sum M_i' = M_i'/M_0'$. وإذا استُخدمت الـ M_i'، نشكل عمودًا للمجاميع المتجمعة T_i للمقادير nM_i' . ومن هذا العمود، وضمن الفترة 1 إلى nM_0' ، نخصص للوحدة i مدى يساوي nM_i' . ولاختيار عيّنة، نسحب عددًا عشوائيًا r بين 1 و M_0' ثم نختار الوحدات الـ n التي تتضمن الأمداء المخصص لها الأعداد $r, r+M_0', \ldots, r+(n-1)M_0'$.

وهذه الطريقة موضحة بالبيان الإحصائي المستخدم في الفقرة (١٩ ـ ٢)، حيث $N=7$. ونفترض أن $n=3$. فبعد تشكيل الـ T_i وتخصيص مدى لكل وحدة، نسحب عددًا عشوائيًا بين 1 و $M_0'=30$ ، ولنقُل مثلاً، $r=17$. فتكون الوحدات المختارة هي تلك التي خُصصت لها أمداء تتضمن الأعداد 77,47,17 ، أي الوحدات 6,4,3 .

وإذا كان $nz_i < 1$ (أي $nM_i' \leq M_0'$) لجميع قيم i ، فإن احتمال اختيار أي وحدة هو nz_i ، ولا يتم اختيار وحدة أكثر من مرة واحدة. وعلى سبيل المثال، نختار الوحدة 5 إذا كان $4 \leqslant r \leqslant 15$ مما يمنح احتمالاً $3z_5=12/30$ وإذا كان $nz_i > 1$ في وحدة

الوحدة	الحجم M'_i	$T_i = 3\sum M_i'$	المدى المخصص	الوحدات المختارة:
1	3	9	1–9	
2	1	12	10–12	
3	11	45	13–45	17 (الوحدة 3)
4	6	63	46–63	47 (الوحدة 4)
5	4	75	64–75	
6	2	81	76–81	77 (الوحدة 6)
7	3	90	82–90	
	—	—		
$M_0' =$	30			

أو أكثر من الوحدات، فيمكن اختيار وحدات كهذه أكثر من مرة في العيّنة، إلا أن متوسط تكرار الاختيار هو $\pi_i = nz_i$. ويحدث هذا في مثالنا هنا في الوحدة 3 ، باعتبار أن $3M_3' = 33 > 30 = M_0'$. ونختـــار الوحدة 3 مرة واحـــدة إذا كان $1 \leqslant r \leqslant 12$ أو $16 \leqslant r \leqslant 30$ (في جميع الاختيارات الـ 27)، ومــرتـين إذا كان $13 \leqslant r \leqslant 15$ (ثـلاثـة اختيارات). ومتوسط تكرار الاختيار هو $3z_3 = 33/30 = (1\times27+2\times3)/30$ ونستنتج أن،

$$\hat{Y}_{sys} = \sum_i^n \frac{y_i}{\pi_i} = \frac{1}{n}\sum_i^n \frac{y_i}{z_i} \qquad (9A.61)$$

تقدير غير منحاز لِـ Y .

وقد اختبر Hartley و Rao (1962) هذه الطريقة مع ترتيب الوحدات أولاً وفق ترتيب عشوائي. وتحت القيد $nz_i<1$ ، (لجميع قيم i)، حصلا على عبارات تقريبية لِـ $v(\hat{Y}_{sys})$ و $V(\hat{Y}_{sys})$.

(١٩ـ١١) طريقة كوكران ـ هارتلي ـ راو (Cochran و Hartley , Rao)

في عيّنـة حجمهـا n ، تشكـل هذه الـطريقـة أولاً n من الزمر العشوائية من الوحدات، لكي تُسحب وحدة واحدة من كل زمرة. ويمكن أن نختار سلفًا أعداد الوحدات $N_1, N_2, \ldots, N_n$ في الزمر المتتالية: وسنرى أنه من المفيد جعل الـ Z_g متساوية قدر الإمكان. وإذا كان Z_g القياس الكلي للحجم في الزمرة g فتُعطى الوحدة i في الزمرة احـتـمـال اخـتـيار z_i/Z_g . وتـقـديـر مجمـوع المـجـتـمـع، Hartley, Rao و Cochran (1962) هو،

$$\hat{Y}_{RHC} = \sum_{g=1}^{n} Z_g \frac{y_g}{z_g} = \sum_{g=1}^{n} \hat{Y}_g \qquad (9A.62)$$

حيث يشير y_g و z_g إلى الوحدة المسحوبة من الزمرة g .

ربما أن الـ Z_g سوف لا تكون متساوية، فسوف لا تحافظ هذه الطريقة على احتمالات اختيار متناسبة مع الحجوم، وهناك بعض الأدلة على أن المقدر في هذه الطريقة يعاني من نقص طفيف في الدقة. وفوائده تكمن في بساطته وعموميته.

وعند تطوير $V(\hat{Y}_{RHC})$ نأخذ المتوسط فوق مرحلتين. المرحلة الأولى هي ترتيب الزمر عشوائيًا، والمرحلة الثانية اختيار وحدة ضمن كل زمرة. ومع أي تقسيم محدّد إلى زمر يكون $\hat{Y}_g$ في (9A.62) مقدّر غير منحاز لمجموع الزمرة Y_g، وبالتالي $E_2(\hat{Y}_{RHC}) = Y$. والعلاقة المعروفة جيدًا لإيجاد التباين فوق مرحلتين من المعاينة، والمبرهنة في الفصل العاشر، هي،

$$V(\hat{Y}_{RHC}) = E_1[V_2(\hat{Y}_{RHC})] + V_1[E_2(\hat{Y}_{RHC})] \qquad (10.2)$$

وبما أن $E_2(\hat{Y}_{RHC}) = Y$ وأنه ثابت، فتباينه صفر، ويختفي الحد الثاني من (10.2) في هذا التطبيق.

ومن أجل $V_2(\hat{Y}_{RHC})$ يمكن أن نستخدم، ضمن زمرة، علاقة التباين. الخاصة بمعاينة مع الإعادة، باعتبار أن وحدة واحدة فقط قد اختيرت من كل زمرة. واحتمالات الاختيار ضمن زمرة هي z_i/Z_g . وبالاستناد إلى (9A.51) في الفقرة (١٩ ـ ٣) نحصل، في أي تقسيم محدد (مع $n_g = 1$)، على،

$$V(\hat{Y}_g) = \sum_{i<j}^{N_g} z_i z_j \left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 \qquad (9A.63)$$

حيث يمتد المجموع فوق الأزواج من الوحدات في الزمرة g . وفوق مجموعة التقسيمات العشوائية إلى زمر يكون احتمال وقوع أي زوج من الوحدات في زمرة g هو $N_g(N_g-1)/N(N-1)$. وبالتالي تكون القيمة المتوسطة لـ $V_2(\hat{Y}_g)$ فوق مجموعة التقسيمات العشوائية هي:

$$E_1 V_2(\hat{Y}_g) = \frac{N_g(N_g-1)}{N(N-1)} \sum_{i}^{N} \sum_{j>i}^{N} z_i z_j \left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 = \frac{N_g(N_g-1)}{N(N-1)} \left(\sum_{i=1}^{N} \frac{y_i^2}{z_i} - Y^2\right) \qquad (9A.64)$$

وذلك بالاستفادة من (9A.63) وبما أن $\hat{Y}_{RHC} = \sum \hat{Y}_g$ ، نجد،

$$E_1[V_2(\hat{Y}_{RHC})] = \frac{\left(\sum_{g=1}^{N} N_g^2 - N\right)}{N(N-1)} \left(\sum_1^N \frac{y_i^2}{z_i} - Y^2\right) = \frac{n\left(\sum_{g=1}^{N} N_g^2 - N\right)}{N(N-1)} V(\hat{Y}_{ppz}) \quad (9A.65)$$

وهكذا يكون $V(\hat{Y}_{RHC})$ ببساطة مساويًا لتباين التقدير $\hat{Y}_{ppz}$ من معاينة مع الإعادة مضروبًا بعامل ما. وإذا كان N/n صحيحًا فالاختيار $N_g = N/n$ يجعل العامل الذي نضرب به أصغر ما يمكن. وفي هذه الحالة

$$V(\hat{Y}_{RHC}) = \left(1 - \frac{n-1}{N-1}\right) V(\hat{Y}_{ppz}) \quad (9A.66)$$

وإذا كان $N = nR + k$ حيث R عدد صحيح و$k < n$ ، فأفضل اختيار هو أخذ k زمرة حجم كل منها $(R+1)$، وحجم كل من الـ $(n-k)$ زمرة الباقية R . وهذا يعطي،

$$V(\hat{Y}_{RHC}) = \left[1 - \frac{n-1}{N-1} + \frac{k(n-k)}{N(N-1)}\right] V(\hat{Y}_{ppz}) \quad (9A.67)$$

وإذا كان N/n صحيحًا فإن (9A.66) تعطي $V(\hat{Y}_{RHC})/V(\hat{Y}_{ppz}) = (N-n)/(N-1)$ ، وهي النسبة نفسها التي نحصل عليها في معاينة عشوائية بسيطة.

ويمكن البرهان على أن،

$$v(\hat{Y}_{RHC}) = \frac{\left(\sum_{g=1}^{n} N_g^2 - N\right)}{\left(N^2 - \sum_{g=1}^{n} N_g^2\right)} \sum_{g=1}^{n} Z_g \left(\frac{y_g}{z_g} - \hat{Y}_{RHC}\right)^2 \quad (9A.68)$$

هو مقدّر غير منحاز للتباين ومع $N = nR + k$، وk من الزمر التي حجمها $(R+1)$، تصبح العلاقة (9A.68) ،

$$v(\hat{Y}_{RHC}) = \frac{N^2 + k(n-k) - Nn}{N^2(n-1) - k(n-k)} \sum_{g}^{n} Z_g \left(\frac{y_g}{z_g} - \hat{Y}_{RHC}\right)^2 \quad (9A.69)$$

(١٩-١٢) مقارنات عددية

في الأدبيات الإحصائية نجد مقارنات لإنجازات بعض الطرق، وبصورة خاصة تلك التي قام بها Rao و Bayless (1970, 1969) في ثلاث حالات: (١) في مجتمعات

اصطناعية صغيرة [مثلاً، المجتمعات التي وضعها Yates و Grundy (1953) ، حيث $N = 4, n = 2$، (ب) تحت نموذج الانحدار الخطي المستخدم في الفقرة (١٩ ـ ٥)، و(ج) في 20 من المجتمعات الفعلية.

وقد قورنت هنا سبع طرق في ثلاثة مجتمعات اصطناعية حيث $N = 5, n = 2$ والحجوم النسبية للوحدات z_i هي نفسها في المجتمعات الثلاثة (A, B, C) جميعها (جدول ١٩ ـ ١). وفي A لا يوجد ارتباط بين المتوسط لكل عنصر، وهو متناسب مع y_i/z_i ، وبين z_i . وفي B يتزايد المتوسط لكل عنصر مع ازدياد الحجوم. وفي C يوجد القليل من الصلة بين مجاميع الوحدات والحجوم.

جدول (١٩ ـ ١) ثلاثة مجتمعات إحصائية صغيرة

الحجوم النسبية	(z_i)	0.1	0.1	0.2	0.3	0.3
المجتمع A	y_i	0.3	0.5	0.8	0.9	1.5
	y_i/z_i	3	5	4	3	5
المجتمع B	y_i	0.3	0.3	0.8	1.5	1.5
	y_i/z_i	3	3	4	5	5
المجتمع C	y_i	0.7	0.6	0.4	0.9	0.6
	y_i/z_i	7	6	2	3	2

والخطط التي قورنت هي كما يلي:

١ ـ معاينة عشوائية بسيطة للوحدات (احتمالات متساوية). التقدير: $\hat{Y}_{SRS} = N\bar{y}$

٢ ـ معاينة عشوائية بسيطة للوحدات. التقدير: نسبة إلى الحجم، $\hat{Y}_R = \sum y_i / \sum z_i$

٣ ـ معاينة باحتمالات متناسبة مع الحجم مع الإعادة. التقدير: $\hat{Y}_{ppz} = (1/n)\sum(y_i/z_i)$

٤ ـ طريقة Brewer ، معاينة بدون إعادة. التقدير: $\hat{Y}_B = (1/n)\sum(y_i/z_i)$

٥ ـ طريقة Murthy . التقدير: $\hat{Y}_M = \left[(1-z_j)\dfrac{y_i}{z_i} + (1-z_i)\dfrac{y_j}{z_j}\right] \Big/ (2 - z_i - z_j)$.

٦ ـ طريقة الـ RHC بزمرة واحدة من ثلاث وحدات، وزمرة بوحدتين. التقدير:

$$\hat{Y}_{RHC} = \sum Z_g (y_g/z_g)$$

٧ ـ طريقة Madow النمطية، بوحدات مرتبة وفق القيم التصاعدة لـ y_i/z_i .

التقدير: $\hat{Y}_{sys} = (1/n)\sum(y_i/z_i)$

وللحصول على فكرة تقريبية عما يمكن توقعه في مجال الإنجازات النسبية لطرق مثل M,B و RHC في مقابل المعاينة العشوائية البسيطة باحتمالات متساوية، نلاحظ أن التباينات النسبية للطرق الجديدة إلى $V(\hat{Y}_{SRS})$ هي على وجه التقريب معاملات الاختلاف النسبية للمقـدارين y_i/z_i و y_i . والنسبـة $cv(y_i/z_i)/cv(y_i)$ ، هي 0.19 في المجتمـع A ، 0.11 في B ، و 6.7 في C . وهكذا ينبغي أن يكون لطرق الاحتمالات غير المتساوية فعالية نسبية تبلغ حوالي 5 في A ، 9 في B ، ولكن 0.15 فقط في C . وهذه المقارنة هي إلى حد ما غير عادلـة بالنسبة لطرق الاحتمال غير المتساوي التي ترجّح مربعات الخطأ في y_i/z_i بقياسات الحجم. وينبغي أن تكون متوسطات مربعات الخطأ لـ $\hat{Y}_R$ قريبة تمامًا من تباينات طرق الاحتمال غير المتساوي.

وفي اختيار طريقة في التطبيقات العملية تكون الاعتبارات الرئيسة هي السهولة التي يمكن بها سحب العيّنـة، إلى جانب بسـاطـة التقـدير ودقتـه وتـوافـر مقـدّر لتبـاينـه. وفي حالة $n=2$ تكون جميع الطرق بسيطة. وفيما يتعلق بالدقة فقد عُرضت الطريقة النمطية هنا في حالتها الأكثر ملاءمة فقط، وهي حالة يندر أن تتوافر للمعاين معلومات تسمح له باستخـدامهـا. وتقدّم جميع الطرق باستثناء $\hat{Y}_{sy}$ و $\hat{Y}_R$ تقديرات عيّنة غير منحازة لتبـاين الخـطأ. ويقـدم الجـدول (١٩ ـ ٢) التبـاينـات (مـع متـوسط مربعـات الخـطأ بالنسبة لـ $\hat{Y}_R$).

جدول (١٩ ـ ٢) تباينات المجاميع المقدّرة للمجتمعات

التقدير	المجتمع A	B	C
$\hat{Y}_{SRS}$	1.575	2.715	0.248
$\hat{Y}_R(MSE)$	0.344	0.351	1.421
$\hat{Y}_{ppz}$	0.400	0.320	1.480
$\hat{Y}_B$	0.246	0.248	1.251
$\hat{Y}_M$	0.267	0.237	1.130
$\hat{Y}_{RHC}$	0.320	0.256	1.184
$\hat{Y}_{sys}$	0.150	0.140	0.760

وفي الاختيار متساوي الاحتمال، تكون نسبة $V(\hat{Y})$ في معاينة بدون إعادة إلى $V(\hat{Y})$ في معاينة مع الإعادة هي $(N\text{-}n)/(N\text{-}1)=0.75$ ، من أجل هذه المجتمعات. ومن أجل $\hat{Y}_{RHC}$ ، نجد أن نسبة $V(\hat{Y}_{RHC})/V(\hat{Y}_{ppz})$ ، حيث يرمز $\hat{Y}_{ppz}$ إلى المعاينة مع الإعادة، هي 0.8 في الجدول (١٩ ـ ٢). ومن أجل الطرق الأخرى في الاحتمال غير المتساوي، تتغير هذه النسبة من مجتمع إلى مجتمع. ومتوسط النسب الثلاث في A ، B و C هو 0.74 في طريقة Brewer ، و 0.73 في طريقة Murthy ، وهي النسبة نفسها تقريبًا كما في حالة طرق الاحتمال المتساوي. والنتائج في حالة $n=3.4$ ، فوق مجتمعات فعلية صغيرة، والتي حصل عليها Bayless و Rao (1970) تتفق بصورة عامة مع $(N-n)/(N-1)$.

وفي المجتمعين A و B نجد أن جميع الطرق الأخرى أكثر دقة بكثير من في معاينة عشوائية بسيطة. وتقدير النسبة إلى الحجم مع معاينة عشوائية بسيطة تنجز تقريبًا إنجازًا مماثلاً لطرق الاحتمال غير المتساوي، مع أنها لا تبلغها تمامًا. وفي هذه الأخيرة، يوجد القليل من الاختيار بين طرق Murthy, Brewer أو RHC . وتنجز الطريقة النمطية في أفضل حالاتها إنجازًا حسنًا جدًّا. وفي المجتمع C ، حيث يوجد القليل من الصلة بين مجاميع الوحدات والحجوم، نجد أن المعاينة العشوائية البسيطة باحتمالات متساوية هي الأفضل بكثير. وكما سنذكر في نهاية الفقرة (١٩ ـ ١٣)، فمن المحتمل أن يعود هذا التفوق إلى المقدّر $N\bar{y}$ وليس إلى ميزة المعاينة العشوائية البسيطة.

وقد قارن Rao و Bayless (1969) 10 طرق احتمالات غير متساوية في 20 مجتمع فعلي عثرا عليها في كتب وأبحاث المعاينة، وحيث تمتد N من 9 إلى 35 وقد اقتصرت دراستهما على طرق (ا) يُعرف أن لها تباينات أصغر من تباين، و(ب) تقدّم مقدّرًا موجبًا وغير منحاز للتباين. ومن بين الطرق التي قُدّمت هنا قارنا فعاليات $\hat{Y}_{RHC}$ ، $\hat{Y}_M$ و $\hat{Y}_{ppz}$ مع فعالية $\hat{Y}_B$. وفي حالة $n=2$ كان هناك القليل مما نختاره بين الطرق الثلاث لمعاينة بدون إعادة، مع سبق طفيف لـ $\hat{Y}_M$ الذي يهزم $\hat{Y}_{RHC}$ حيثما يختلف التقديران في الدقة. وقد لاحظنا أيضًا (فقرة ١٩ ـ ٧) أن مقدّر التباين يمكن أن يكون غير مستقر في طرق تستخدم تقدير Horvitz-Thompson . وتقارن نتائج Rao-Bayless

Yates-Grundy في شكل $v(\hat{Y}_B)$ و $v(\hat{Y}_{RHC})$، $v(\hat{Y}_M)$ مع معامل معامل اختلاف $v(\hat{Y}_B)$، وذلك كمقاييس لاستقرار مقدرات التباين. وبالنسبة إلى اختلاف $v(\hat{Y}_{ppz})$ معتبرًا 100% ، كانت الفعاليات الوسطى لمقدّرات التباين الأخرى: $v(\hat{Y}_{RHC})=$ 109% ، $v(\hat{Y}_M)=$ 104% ، $v(\hat{Y}_{PPZ})=$ 97% ، وتُظهر الطرق الثلاث جميعها القليل من المكاسب الفردية الكبيرة.

ويعطي Bayless و Rao (1970) مقارنات مشابهة في حالة $n=3$ (14 مجتمعًا) و $n=4$ (10 مجتمعات). وقد استُخدم توسيع Sampford لطريقة Brewer وفي حالة n أكبر بكثير من 2 تستدعي كل من طريقتي Sampford و Murthy مساعدة الحاسب عند حساب الاحتمالات المطلوبة. وتباينا $\hat{Y}_M$ و $\hat{Y}_S$ قريبان جدًّا من بعضهما في جميع المجتمعات تقريبًا، ويتأخر $\hat{Y}_{RHC}$ عنهما قليلاً، وتهبط فعاليته الوسطى بالمقارنة مع $\hat{Y}_S$ إلى 92% في حالة $n=4$. وعلى الوجه الآخر، وفيما يتعلق باستقرار مقدّرات التباين، يزداد تفوق $\hat{Y}_{RHC}$ و $\hat{Y}_M$ فالفعاليات الوسطى النسبية هي 118% $(n=3)$ و 129% $(n=4)$ من أجل $v(\hat{Y}_{RHC})$ ، و 110% $(n=3)$ و 120% $(n=4)$ من أجل $v(\hat{Y}_M)$ وذلك بالنسبة إلى 100% من أجل $v(\hat{Y}_S)$.

وقارن Rao و Bayless أيضًا فعاليتي $\hat{Y}$ و $v(\hat{Y})$ لبعض المقدّرات تحت نموذج الانحدار الخطي في الفقرة (١٩ ـ ٥) مع $\alpha=0$ وبينما تعتمد نتائج المقارنة على القوة g ، كان الاتجاه العام مماثلاً لما رأيناه في المجتمعات الفعلية.

(١٩ ـ ١٣) التقديرات النسبة والتقديرات الطبقية

قُدمت العلاقات السابقة باعتبارها تنطبق على طبقة بمفردها، مع أن التركيز على قيمة صغيرة لِـ n تتضمن وجود تقسيم طبقي سابق وفقًا لمبدأ آخر (مثلاً، الموقع الجغرافي، مدينة ـ ريف). وتمديد العلاقات إلى هذا التقسيم إلى طبقات يتمّ كالمعتاد. ففي أي طريقة، إذا كان $n_h, z_{hi}, \pi_{hi}, y_{hi}$ وهلمّ جرًّا، يشير إلى الطبقة h ، فعندئذ،

$$\hat{Y}=\sum_h^L \hat{Y}_h:\ V(\hat{Y})=\sum_h^L V(\hat{Y}_h):\ v(\hat{Y})=\sum v(\hat{Y}_h) \qquad (9A.70)$$

وتدخل التقديرات النسبة في الاعتبار إما عندما يكون المتغير المدروس نسبة (مثلاً، عدد النساء العاطلات عن العمل مقسومًا على عدد النساء القادرات على

العمل)، أو عندما تستخدم لزيادة الدقة. وفي المعاينة باحتمالات غير متساوية يجب القيام باختيار واحد لِـ z_i أو z_{hi} من أجل جميع المتغيرات التي نريد تقديرها. وعلى سبيل المثال، قد تكون الـ z_i أو الـ z_{hi} متناسبة مع المبيعات الإجمالية الأخيرة لنوع من النشاط التجاري، بينما يهدف المسح الإحصائي لتقدير المبيعات الجارية لأصناف معيّنة من المفردات بالإضافة إلى المبيعات الإجمالية الجارية. وفي بعض الأصناف قد لا تكون المبيعات متناسبة تمامًا مع الـ z_i. وفي حالات كهذه قد يؤدي استخدام التقدير المألوف «النسبة إلى آخر قيمة افترضها المتغير نفسه» إلى زيادة كبيرة في الدقة.

وفي حالة معاينة باحتمالات غير متساوية يمكن بسهولة القيام بتغيير في العلاقات بحيث تلائم المقدّرات النسبة. وفي مجتمع غير مقسم إلى طبقات نضع، من أجل أي طريقة، $X\hat{Y}/\hat{X}$ بدلاً من $\hat{Y}$ وعلى سبيل المثال، نستخدم في حالة المقدّر HT، $X(\sum^n y_i/\pi_i)/\sum^n (x_i/\pi_i)$ كالمقدّر النسبة بدلاً من $\sum (y_i/\pi_i)$. وللتقريبات المعتادة لمتوسط مربعات الخطأ للتقدير النسبة وكذلك لتقدير متوسط المربعات الخطأ هذا، نضع في عبارة $V(\hat{Y})$ المقدار $d_i = (y_i - Rx_i)$ بدلاً من y_i وفي $v(\hat{Y})$ نضع $d_i' = (y_i - \hat{R}x_i)$ بدلاً من y_i.

وعلى سبيل المثال، في حالة مقدّر Horvitz-Thompson على شكل نسبة نجد من المعادلة (9A.44) أن العبارة التقريبية لِـ v_2، (وهو تقدير التباين وفق صيغة Yates-Grundy) هي،

$$v_2[\hat{Y}_{HT(R)}] \doteq \sum_{i}^{N} \sum_{j>i}^{N} \frac{(\pi_i\pi_j - \pi_{ij})}{\pi_{ij}} \left(\frac{d_i'}{\pi_i} - \frac{d_j'}{\pi_j}\right)^2 \qquad (9A.71)$$

ومن المرجّح، في مجتمع مقسّم إلى طبقات مع n_h صغير، استخدام التقدير النسبة المركبّ (فقرة ٦ - ١١) [أي، $[\hat{Y} = X(\sum \hat{Y}_h)/(\sum \hat{X}_h)]$. ومن أجل علاقات تباين تقريبية، نضع في V المتغير $d_{hi} = y_{hi} - Rx_{hi}$ بدلاً من y_{hi}، وفي v نضع $d_{hi}' = y_{hi} - \hat{R}x_{hi}$ بدلاً من y_{hi}. وعندما لا تكون الـ y_i لمفردة على صلة بالـ z_i ولا يتوافر التقدير النسبة الملائم لمفردة كهذه، فقد درس Rao (1966) مقدّرات بديلة. ويمكن توليد هذه البدائل من أي مقدّرات باحتمالات غير متساوية (RHC, M, TH الخ) وذلك بوضع Ny_i بدلاً من $\frac{y_i}{z_i}$ حيثما يظهر y_i في عبارة المقدّر. وهكذا تكون الصيغة البديلة لمقدّر Horvitz-Thompson هي،

$$\hat{Y}^*_{HT} = \frac{N}{n}\sum_i^n y_i \quad (9A.72)$$

بينما تكون في طريقة RHC ،

$$\hat{Y}^*_{RHC} = N\sum_g^n Z_g y_g \quad (9A.73)$$

والمقدّرات منحازة إلا أن البداهة تقترح أنه إذا لم يكن لِـ y_i صلة بِـ z_i فينبغي أن تكون الانحيازات صغيرة نسبيًا. وبالطريقة نفسها المستخدمة في إيجاد $V(\hat{Y}_{HT})$ في (9A.42) نجد،

$$V[\hat{Y}^*_{HT}] = \frac{N^2}{n^2}\sum_i^N \sum_{j>i}^N (\pi_i\pi_j - \pi_{ij})(y_i - y_j)^2 \quad (9A.74)$$

وهكذا يعتمد $V[\hat{Y}^*_{HT}]$ على مقدار التغير في مجاميع الوحدات، كما في حالة $V(\hat{Y}_{SRS})$ وفي المجتمع C ، جدول (١٩ ـ ٢)، أعطت هذه الطريقة 0.266 و 0.28:0 من أجل متوسطات مربعات الخطأ للصيغتين البديلتين لِـ $\hat{Y}_{HT}$ و $\hat{Y}_M$ ، وهاتان الصيغتان هما تقريبًا في نفس مستوى جودة $\hat{Y}_{SRS}$. وفي الطريقتين كلتيهما كان الحد الذي يحوي مربع الانحياز حوالي 4% من متوسط مربعات الخطأ.

تمارين

(١٩ ـ ١) يعطي Horvitz و Thompson (1952) البيان الإحصائي التالي لتقديرات بالعين المجرّدة M_i ، وللأعداد الفعلية y_i للمنازل في جادّات مدينة آميس في ولاية أيووا. وللمساعدة في الحسابات، أعطيت أيضًا قيم $\bar{y}_i$ و $\bar{y}_i^2/M_i$. وقد اختيرت عيّنة من جادّة واحدة. احسب تباينات العدد الكلي للمنازل Y ، كما نحصل عليه من

M_i	y_i	$\bar{y}_i$	y_i^2/M_i	M_i	y_i	$\bar{y}_i$	y_i^2/M_i
9	9	1.0000	9.000	19	19	1.0000	19.000
9	13	1.4444	18.778	21	25	1.1905	29.762
12	12	1.0000	12.000	23	27	1.1739	31.696
12	12	1.0000	12.000	24	21	0.8750	18.375
12	14	1.1667	16.333	24	35	1.4583	51.042
14	17	1.2143	20.643	25	22	0.8800	19.360
14	15	1.0714	16.071	26	25	0.9615	24.038
17	20	1.1765	23.529	27	27	1.0000	27.000
18	19	1.0556	20.056	30	47	1.5667	73.633
18	18	1.0000	18.000	40	37	0.9250	34.225

(ا) التقدير غير المنحاز في معاينة باحتمالات متساوية ، (ب) التقدير النسبة باحتمالات متساوية ، (ج) معاينة باحتمال متناسب مع M_i (في حالة التقدير النسبة ، أحسب متوسط مربعات الخطأ الفعلي وليس العلاقة التقريبية). هل تتفق النتائج مع المناقشة في الفقرة (١٩ ـ ٥)؟

(١٩ ـ ٢) أرسل استبيان إلى عيّنة من المدارس الثانوية للتعرّف على المدارس التي تقدّم تسهيلات أو خدمات معيّنة ، مثلاً ، مقررًا في اللغة الروسية أو في السباحة . إذا كان عدد الطلاب في المدرسة i هو M_i فالكمية المراد تقديرها من أجل خدمة معيّنة هي P نسبة طلاب الثانوي الذين يتلقون هذه الخدمة في المدارس ، أي

$$P = \frac{\sum_{w} M_i}{\sum_{i=1}^{N} M_i}$$

حيث $\sum_{w}$ هو المجموع فوق المدارس التي تتوافر فيها هذه الخدمة .

سحبنا عيّنة من n مدرسة باحتمال متناسب مع M_i مع الإعادة . ومن أجل خدمة معيّنة وُجد أن a مدرسة من بين n من المدارس تقدم هذه الخدمة (ا) بيّن أن $\hat{P} = \frac{a}{n}$ تقدير غير منحاز لِـ P وأن تباينه الصحيح هو $P(1-P)/n$ (تلميح : في نتيجة النظرية (٩ ـ ٤) ، ليكن $y_i = M_i$ إذا توافرت هذه الخدمة في المدرسة و 0 فيما عدا ذلك .) (ب) بيّن أن $v(\hat{P}) = \hat{P}(1-\hat{P})/(n-1)$ تقدير غير منحاز لِـ $V(\hat{P})$.

(١٩ ـ ٣) ترتّب الوحدات الكبيرة في مجتمع نفسها في فصول من الحجوم عددها منتـه . وكل وحدة من الفصل h تتضمن M_h من الوحدات الصغيرة . (ا) تحت أية شروط تعطي معاينة الـ أ م ح ، في المتوسط ، توزيع تكرار لفصول الحجم ضمن العيّنة مطابق للتوزيع الذي يعطيه التقسيم إلى طبقات وفقًا للحجم مع محاصّة مثلى وحجم مثبت للعيّنـة؟ (ب) إذا كان التبـاين بين الـوحـدات الكبيرة في الفصل h هو kM_h حيث k ثابت في جميع الفصول ، فما هو نظام احتمالات اختيار الوحدات الذي يعطي عيّنة يكون للحجوم فيها ، تقريبًا ، نفس التوزيع الذي نجده في عيّنة عشوائية طبقية مع محاصّة مثلى وحجم مثبت للعيّنة؟

(١٩-٤) إذا كان $N=3$ في مجتمع $z_i=\frac{1}{2},\frac{1}{3},\frac{1}{6}$ و $y_i=7,5,2$ وسحبنا وحدتين بدون إعادة، الأولى باحتمال متناسب مع z_i والثانية باحتمال متناسب مع الحجوم الباقية. (ا) تحقق من أن $\pi_1=\frac{51}{60}$، $\pi_2=\frac{44}{60}$، $\pi_3=\frac{25}{60}$ وأن $\pi_{12}=\frac{35}{60}$ $\pi_{13}=\frac{16}{60}$ $\pi_{23}=\frac{9}{60}$. (ب) من أجل هذه الطريقة في اختيار العيّنة، قارن تبايني $\hat{Y}_{HT}$ و $\hat{Y}_M$ ثم قارنهما أيضًا مع تباين المقدّر $\hat{Y}_{RHC}$ مستخدمًا طريقته في اختيار العيّنة. يمكنك إما حساب كل التقديرات الثلاثة الممكنة أو استخدام علاقات التباين. (ج) بيّن أن نسبة $V(\hat{Y}_M)$ إلى $V(\hat{Y}_{ppz})$ في معاينة مع الإعادة قريبة من القيمة 1/2 التي تُطبّق في معاينة باحتمالات متساوية.

(١٩ - ٥) من أجل المجتمع في التمرين (١٩ - ٤) يأخذ متغير ثان القيم $y_{2i}=8,5,9$، التي ليس لها صلة بالمقادير z_i بحيث يُتوقّع أن يكون أداء $\hat{Y}_{HT}$ و $\hat{Y}$ من أجل هذا المتغير رديئًا عند استخدام طريقة المعاينة المذكورة في التمرين (١٩-٤). قارن متوسط مربعات الخطأ للتقدير $\hat{Y}^*_{HT}=1.5(y_{2i}+y_{2j})$ وهو تقدير (Rao) مع تباين $\hat{Y}_{SRS}=1.5(y_{2i}+y_{2j})$ في معاينة باحتمالات متساوية. ما مقدار مساهمة الانحياز في متوسط مربعات الخطأ؟

(١٩ - ٦) من أجل طريقة Brewer حيث $n=2$، أظهرت الفقرة (١٩ - ٨) أن،

$$\pi_{ij}=\frac{4z_iz_j(1-z_i-z_j)}{(1-2z_i)(1-2z_j)}\Big/\left(1+\sum_{i}^{N}\frac{z_i}{1-2z_i}\right)$$

(ا) بيّن أنه إذا كان كل $z_i<\frac{1}{2}$ فعندئذ، $0<\pi_{ij}<4z_iz_j$ (كل $i\neq j$)

(ب) بيّن أن هذه النتيجة تجعل تباين مقدّر Yates-Grundy موجبًا دومًا في هذه الطريقة. تلميح: في (ا) يكفي تبيان أن،

$$(1-z_i-z_j)=(1-2z_i)(1-2z_j)\left[1+\frac{z_i}{1-2z_i}+\frac{z_j}{1-2z_j}\right]$$

(١٩ - ٧) (ا) في طريقة Durbin مع $n=2$، تحقق مباشرة من أن احتمال سحب الوحدة j في السحب الثاني هو z_j كما عرضنا سابقًا بالصفحة ٣٧٨.

(ب) في حالة $N=4$ ، $z_i=0.1, 0.2, 0.3, 0.4$ ، احسب احتمال سحب الوحدة 1 أولاً واحتمال سحبها ثانيًا وذلك في طريقة Brewer . تحقق أن مجموع الاحتمالين يساوي 0.2 أي $2z_i$.

(١٩ ـ ٨) في طريقة Madow النمطية، يمكن اختيار وحدة إلى العيّنة أكثر من مرة إذا كان $nz_i > 1$ (أي $nM_i' > M_0'$). بيّن (كما عرضنا على الصفحة ٣٨٢) أنـه مع وحـدات كهـذه يكون التكرار المتوسط للاختيار مساويًا لِـnz_i ، بحيث يبقى مقدّر Horvitz-Thompson لِـY غير منحاز في حالة $nz_i > 1$.

الفصل العاشر

المعاينة الجزئية بوحدات متساوية الحجم

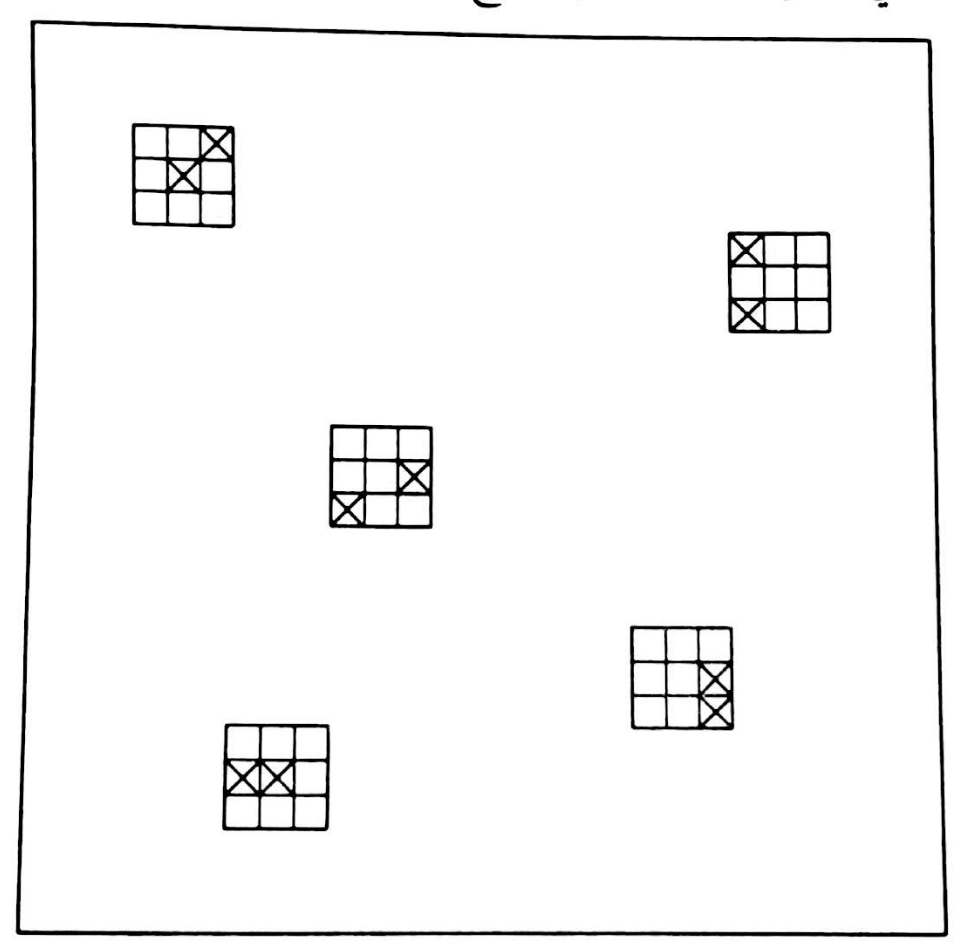
(١٠-١) معاينة على مرحلتين

لنفرض أنه يمكن تقسيم كل وحدة في المجتمع إلى عدد من الوحدات الأصغر أو الوحدات الجزئية. واختيرت عيّنة تحوي n من الوحدات. فإذا كانت الوحدات الجزئية ضمن وحدة مختارة تعطي نتائج متماثلة، فيبدو من غير الاقتصادي أن نقيسها جميعًا. والإجراء الشائع هو اختيار وقياس عيّنة من الوحدات الجزئية في أي وحدة مختارة. وتدعى هذه الطريقة بالمعاينة الجزئية، باعتبار أننا لم نقم بقياس الوحدة بكاملها، وإنما أخذنا عيّنة من الوحدة نفسها. والاسم الآخر الذي يعود إلى Mahalanobis هو المعاينة على مرحلتين، لأن العيّنة تؤخذ في خطوتين. الأولى هي أن نختار عيّنة من الوحدات، تسمى غالبًا الوحدات الأولية، والثانية هي أن نختار عيّنة من وحدات المرحلة الثانية أو الوحدات الفرعية من كل وحدة أولية مختارة.

وللمعاينة الجزئية تطبيقات متنوعة جدًا، تذهب بعيدًا فيها وراء المجال المباشر لنظرية المعاينة الإحصائية. وحيثما تتضمن أية عملية، كيميائية أو فيزيائية أو بيولوجية، اختبارات يمكن إجراؤها على قدر صغير من المادة التجريبية، فمن المرجّح أن تُسحب هذه المادة كعيّنة جزئية من مقدار أكبر هو بحدّ ذاته عيّنة.

وسندرس في هذا الفصل أبسط حالة، تحتوي فيها كل وحدة على العدد M نفسه من الوحدات الجزئية التي نختار من بينها m عنصرًا عند أخذ عيّنة من أي وحدة. ويبيّن الشكل (١٠-١) تمثيلاً تخطيطيًا للمعاينة على مرحلتين، حيث $M=9$ و $m=2$.

والميزة الرئيسة للمعاينة على مرحلتين هي أنها أكثر مرونة من المعاينة على مرحلة واحدة. وتختزل إلى المعاينة على مرحلة واحدة عندما يكون $m=M$ ، ولكن ما لم يكن هذا هو أفضل اختيار لـ m ، فلدينا الفرصة لأخذ قيمة ما أصغر وأكثر فعالية فيما يبدو. وكالمعتاد فإن القضية تختزل إلى مسألة توازن بين الدقة الإحصائية والتكلفة. وعندما تتفق الوحدات الجزئية في الوحدة نفسها اتفاقًا شديدًا فيما بينها، فإن اعتبارات الدقة تقترح قيمة صغيرة لـ m . وعلى الوجه الآخر، تكون تكلفة قياس كامل الوحدة أحيانًا في نفس تكلفة معاينتها جزئيًا، مثلاً، عندما تكون الوحدة أسرة، واستجابة شخص واحد فيها تعطي معلومات دقيقة حول جميع عناصر الأسرة.



☒ يرمز لعنصر من العيّنة

شكل (١٠-١) تمثيل تخطيطي لمعاينة على مرحلتين ($m=2, M=9, n=5, N=81$)

(١٠-٢) إيجاد المتوسطات والتباينات في معاينة على مرحلتين

في المعاينة على مرحلتين تعطي خطة المعاينة أولاً طريقة لاختيار n من الوحدات.

وعندئذ، ولكل وحدة وقع عليها الاختيار، تعطي طريقة لاختيار عدد من الوحدات الجزئية منها. وعند إيجاد متوسط وتباين تقدير، يجب أخذ المتوسط فوق جميع العيّنات التي يمكن توليدها بالطريقة ذات المرحلتين هذه. وإحدى الطرق لحساب هذا المتوسط هو أن نحسب أولاً متوسط التقدير فوق جميع اختيارات المرحلة الثانية التي يمكن سحبها من مجموعة مثبتة من n من الوحدات التي تختارها الخطة. ثم نحسب المتوسط فوق جميع الاختيارات الممكنة وفق الخطة لِـ n من الوحدات. ولتقدير $\hat{\theta}$ يمكن التعبير عن هذه الطريقة على الشكل،

$$E(\hat{\theta}) = E_1[E_2(\hat{\theta})] \qquad (10.1)$$

حيث يرمز E_2 لقيمة التوقع أو المتوسط فوق جميع العيّنات، ويرمز لعملية أخذ المتوسط فوق جميع اختيارات المرحلة الثانية الممكنة من مجموعة مثبتة من الوحدات، ويرمز E_1 لعملية أخذ المتوسط فوق جميع اختيارات المرحلة الأولى.

ومن أجل $V(\hat{\theta})$ تعطي هذه الطريقة النتيجة التالية التي يمكن تذكّرها بسهولة.

$$V(\hat{\theta}) = V_1[E_2(\hat{\theta})] + E_1[V_2(\hat{\theta})] \qquad (10.2)$$

حيث $V_2(\hat{\theta})$ هو التباين فوق جميع اختيارات العيّنة الجزئية الممكنة من أجل مجموعة معطاة من الوحدات. ولتبيان هذا، ليكن $\theta = E(\hat{\theta})$ (حيث لا يمثل θ بالضرورة المقدار الذي يُراد لِـ $\hat{\theta}$ أن يقدّره، باعتبار أن $\hat{\theta}$ يمكن أن يكون منحازًا). وبالتعريف

$$V(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = E_1E_2(\hat{\theta} - \theta)^2 \qquad (10.3)$$

ولكن

$$E_2(\hat{\theta} - \theta)^2 = E_2(\hat{\theta}^2) - 2\theta E_2(\hat{\theta}) + \theta^2 \qquad (10.4)$$

$$= [E_2(\hat{\theta})]^2 + V_2(\hat{\theta}) - 2\theta E_2(\hat{\theta}) + \theta^2 \qquad (10.5)$$

لنأخذ الآن المتوسط فوق اختيارات المرحلة الأولى. فبما أن $E_1E_2(\hat{\theta}) = \theta$، نجد،

$$V(\hat{\theta}) = E_1[E_2(\hat{\theta})]^2 - \theta^2 + E_1[V_2(\hat{\theta})] \qquad (10.6)$$

$$= V_1[E_2(\hat{\theta})] + E_1[V_2(\hat{\theta})] \qquad (10.7)$$

ويمكن تعميم العلاقة (10.7) بصورة طبيعية إلى ثلاث مراحل أو أكثر. وفي حالة معاينة على ثلاث مراحل نجد،

$$V(\hat{\theta}) = V_1\{E_2[E_3(\hat{\theta})]\} + E_1\{V_2[E_3(\hat{\theta})]\} + E_1\{E_2[V_3(\hat{\theta})]\} \qquad (10.7')$$

(١٠-٣) تبايـن تقـديـر المتوسط في معاينة على مرحلتين

تُستخدم الرموز التالية :

y_{ij}= القيمة التي نحصل عليها من الوحدة الجزئية j في الوحدة الأولية i .

$\bar{y}_i = \sum_{j=1}^{m} \frac{y_{ij}}{m}$ = متوسط العيّنة لكل وحدة جزئية في الوحدة الأولية i .

$\bar{\bar{y}} = \sum_{i=1}^{n} \frac{\bar{y}_i}{n}$ = متوسط العينة الإجمالي لكل وحدة جزئية .

$S_1{}^2 = \frac{\sum_{i=1}^{N} (\bar{Y}_i - \bar{\bar{Y}})^2}{N-1}$ = التباين بين متوسطات الوحدات الأولية .

$S_2{}^2 = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} (y_{ij} - \bar{Y}_i)^2}{N(M-1)}$ = التباين بين الوحدات الجزئية ضمن الوحدات الأولية .

ونلاحظ أن Y_i يرمز للمجموع الإجمالي للوحدات الجزئية في الوحدة i (وكنا رمزنا له في الفصلين ٩ و١٩ بـِ y_i) .

نظرية (١٠-١)

إذا كانت الـوحـدات الـ m والـوحدات الجزئية الـ n من كل وحدة مختارة قد اختيرت وفق معاينة عشوائية بسيطة، فعندئذ يكون $\bar{\bar{y}}$ تقديرًا غير منحاز لِـ $\bar{\bar{Y}}$ بتباين.

$$V(\bar{\bar{y}}) = \left(\frac{N-n}{N}\right)\frac{S_1{}^2}{n} + \left(\frac{M-m}{M}\right)\frac{S_2{}^2}{mn} \qquad (10.8)$$

برهان

مع معاينة عشوائية بسيطة في المرحلتين كلتيهما نجد،

$$E(\bar{\bar{y}}) = E_1[E_2(\bar{\bar{y}})] = E_1\left(\frac{1}{n}\sum^{n} \bar{Y}_i\right) = \left(\frac{1}{N}\sum^{N} \bar{Y}_i\right) = \bar{\bar{Y}} \qquad (10.9)$$

ومن أجل $V(\bar{\bar{y}})$ نستخدم العلاقة (10.2) .

$$V(\bar{\bar{y}}) = V_1[E_2(\bar{\bar{y}})] + E_1[V_2(\bar{\bar{y}})] \qquad (10.10)$$

وبما أن $E_2(\bar{\bar{y}}) = \sum^{n} \bar{Y}_i/n,$ فالحد الأول على اليمين هو تباين المتوسط لكل وحدة جزئية من عيّنة عشوائية بسيطة تتضمن n وحدة وعلى مرحلة واحدة. وبالتالي نجد من النظرية (٢-٢)،

$$V_1[E_2(\bar{\bar{y}})] = \left(\frac{N-n}{N}\right)\frac{S_1{}^2}{n} \qquad (10.11)$$

وفضلاً عن ذلك، مع $\bar{\bar{y}} = \sum^{n} \bar{y}_i/n$ واستخدام معاينة عشوائية بسيطة في المرحلة الثانية، نجد،

$$V_2(\bar{\bar{y}}) = \frac{(M-m)}{Mn^2}\,\frac{\sum^{n} S_{2i}{}^2}{m} \qquad (10.12)$$

حيث $S_{2i}{}^2 = \sum_{j}^{M} (y_{ij} - \bar{Y}_i)^2/(M-1)$ التبـاين بين الوحدات الجزئية في الوحدة الأولية i . وعندما نأخـذ المتـوسط فوق عيّنـات المـرحلة الأولى، نجـد أن متوسط المقادير $\sum_{i}^{n} S_{2i}{}^2/n$ هو $\sum_{i}^{N} S_{2i}{}^2/N = S_2{}^2$.

وبالتالي،

$$E_1[V_2(\bar{\bar{y}})] = \left(\frac{M-m}{M}\right)\frac{S_2{}^2}{mn} \qquad (10.13)$$

وتنتج النظرية من العلاقة (10.10) لدى جمع (10.11) و (10.13) .

وإذا كان $f_1 = n/N$ و $f_2 = m/M$ كسري المعاينة في المرحلتين الأولى والثانية، فالصيغة البديلة للنتيجة هي،

$$V(\bar{\bar{y}}) = \frac{1-f_1}{n}S_1{}^2 + \frac{1-f_2}{mn}S_2{}^2 \qquad (10.14)$$

(١٠-٤) تقدير عيّنة للتباين

نظرية (١٠-٢): تحت شروط النظرية (١٠-١) يكون

$$v(\bar{\bar{y}}) = \frac{1-f_1}{n} s_1^2 + \frac{f_1(1-f_2)}{mn} s_2^2 \tag{10.15}$$

تقديرًا غير منحاز لـ $V(\bar{\bar{y}})$ حيث $f_1 = n/N$, و $f_2 = m/M$، و

$$s_1^2 = \frac{\sum^n (\bar{y}_i - \bar{\bar{y}})^2}{n-1} \qquad s_2^2 = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2}{n(m-1)} \tag{10.16}$$

برهان

$$(n-1)s_1^2 = \sum^n (\bar{y}_i - \bar{\bar{y}})^2 = \sum^n \bar{y}_i^2 - n\bar{\bar{y}}^2 \tag{10.17}$$

وبالتالي

$$(n-1)E_2(s_1^2) = \sum^n \bar{Y}_i^2 + \sum^n \frac{(1-f_2)}{m} S_{2i}^2 - n\bar{\bar{Y}}_n^2 - \frac{1}{n}\sum^n \frac{(1-f_2)}{m} S_{2i}^2 \tag{10.18}$$

حيث $\bar{\bar{Y}}_n = \sum^n Y_i/n$. ويصحّ الحد الأخير على اليمين باعتبار أن المعاينة الجزئية مستقلة في الوحدات المختلفة و $\bar{\bar{y}} = \sum^n \bar{y}_i/n$ ، وهكذا نجد،

$$(n-1)E_2(s_1^2) = \sum^n (\bar{Y}_i - \bar{\bar{Y}}_n)^2 + \frac{(n-1)(1-f_2)}{nm} \sum^n \frac{S_{2i}^2}{n} \tag{10.19}$$

ولدى الضرب بـ $(1-f_1)/n(n-1)$ وأخذ المتوسط فوق المرحلة الأولى من المعاينة العشوائية البسيطة، نجد،

$$E\frac{(1-f_1)}{n} s_1^2 = \frac{(1-f_1)}{n} S_1^2 + \frac{(1-f_1)(1-f_2)}{mn} S_2^2 \tag{10.20}$$

وبالمقارنة مع (10.14) الخاصة بـ $V(\bar{\bar{y}})$ ، نلاحظ أن الحد المتضمن S_2^2 أصغر بمقدار $f_1(1-f_2)S_2^2/mn$ وبما أن $E_1E_2(s_2^2) = S_2^2$ فالتقدير

$$v(\bar{\bar{y}}) = \frac{1-f_1}{n} s_1^2 + \frac{f_1(1-f_2)}{mn} s_2^2 \tag{10.21}$$

هو إذن تقدير غير منحاز لـ $V(\bar{\bar{y}})$.

نتيجة

نجد من (10.20) نتيجة سنستخدمها فيما بعد هي ،

$$E(s_1^2) = S_1^2 + \frac{(1-f_2)}{m} S_2^2 = S_1^2 + \frac{S_2^2}{m} - \frac{S_2^2}{M} \tag{10.22}$$

ومنه نستنتج أن ،

$$s_1^2 - \frac{s_2^2(1-f_2)}{m}$$

تقدير غير منحاز لِـ S_1^2 .

ملاحظات على النظرية (١٠-٢)

إذا كان $m=M$ ، أي $f_2=1$ فتصبح العلاقة (10.15) العلاقة المناسبة لمعاينة عشوائية بسيطة من الوحدات. وإذا كان $n=N$ ، فالعلاقة هي تلك الخاصة بمعاينة عشوائية طبقية تناسبية ، طالما أنه يمكن اعتبار الوحدات الأولية عندئذ كطبقات، وأن عيّنة أُخذت من كل منها. وفي هذا المجال نقول إن المعاينة على مرحلتين هي نوع من التقسيم غير الكامل إلى طبقات، حيث الطبقات هي الوحدات.

وعندما يكون $f_1=n/N$ مهملاً ، نحصل على النتيجة البسيطة ،

$$v(\bar{y}) = \frac{s_1^2}{n} = \frac{\sum_{i=1}^{n} (\bar{y}_i - \bar{y})^2}{n(n-1)} \tag{10.23}$$

وهكذا يمكن حساب تقدير التباين من معرفة متوسطات الوحدات فقط. وتكون هذه النتيجة مفيدة في حالة معاينة جزئية نمطية، باعتبار أننا لا نستطيع في هذه الحالة حساب تقدير غير منحاز لِـ S_2^2 . إلا أن (10.33) لا تزال ممكنة التطبيق، شريطة أن تكون n/N صغيرة. وإذا لم تكن n/N صغيرة فإن (10.23) تبالغ في تقدير الكمية $f_1 S_1^2/n$ ، كما نرى من (10.20) و (10.14) .

(١٠-٥) تقدير النسب

إذا كانت العناصر مصنفة في فصلين، ونرغب في تقدير نسبة العناصر التي تقع في الفصل الأول، فيمكن تطبيق العلاقات السابقة من خلال التدبير المعتاد الذي نعرّف بموجبه y_{ij} بأنه يساوي 1 إذا وقعت الوحدة الجزئية في هذا الفصل وتساوي

صفـرًا فيما عدا ذلـك. لتكن $p_i = a_i/m$ نسبة العناصر من العيّنة الجزئية من الـوحدة i التي تقع في الفصل الأول. فعندئذ يصبح تقديرا التباين s_1^2 و s_2^2 المطلوبان في النظرية (١٠-٢) كما يلي:

$$s_1^2 = \frac{\sum_{i=1}^{n} (p_i - \bar{p})^2}{n-1}$$

$$s_2^2 = \frac{m}{n(m-1)} \sum_{i=1}^{n} p_i q_i$$

حيث $\bar{p} = \sum p_i / n$ وبالتالي نجد من النظرية (١٠-٢)، ما يلي:

$$v(\bar{p}) = \frac{1-f_1}{n(n-1)} \sum_{i}^{n} (p_i - \bar{p})^2 + \frac{f_1(1-f_2)}{n^2(m-1)} \sum_{i}^{n} p_i q_i \qquad (10.24)$$

مثـــال

في دراسة لمرض نباتي زُرعت الأشجار في 160 قطعة أرض صغيرة تتضمن كل منها 9 شجرات. واختيرت عيّنة عشوائية من 40 قطعة، وثلاث شجرات عشوائية من كل قطعة ضمتها العيّنة، لاختبار وجود المرض فيها. وقد وُجد أن 22 قطعة لم تتضمن أية شجرة مريضة (من بين الثلاث موضع الاختبار) وفي 11 منها وُجدت شجرة واحدة مريضة، كما وُجدت شجرتان في أربع منها، وثلاث في ثلاث منها. قدّر نسبة الأشجار المريضة وخطأها المعياري. ويدل الرمز ϕ على التكرارات 22 ,11 ,4 ,3.

لدينا $N = 160, M = 9, n = 40, m = 3$. ولإيجاد s_1^2 و s_2^2 من المربح أن نعمل أولاً بأعداد الأشجار المريضة $(3p_i)$، وأعداد الأشجار الصحيحة $(3q_i)$. وقد رُتّبت الحسابات كما يلي:

$3p_i$	التكرار ϕ	$9p_iq_i$	$9\phi p_iq_i$	$3\phi p_i$	$9\phi p_i^2$
0	22	0	0	0	0
1	11	2	22	11	11
2	4	2	8	8	16
3	3	0	0	9	27
	40		30	28	54

$$\bar{p}=\frac{3\sum \phi p_i}{3\sum \phi}=\frac{28}{120}=0.233$$

$$\sum\phi(p_i-\bar{p})^2=\frac{1}{(9)}\left(54-\frac{(28)^2}{40}\right)=3.822$$

$$\sum\phi p_i q_i=\frac{30}{9}=3.333$$

وبالتالي نجد من العلاقة السابقة لهذا المثال مباشرة،

$$v(\bar{p})=\frac{(3)(3.822)}{(4)(40)(39)}+\frac{(2)(3.333)}{(4)(3)(1600)(2)}=0.00201$$

ونسبة الأشجار المريضة هي 0.233 بخطأ معياري 0.045 . والعلاقة التقريبية $s_1/\sqrt{n}$ من (10.23) تعطي 0.049 وهو تقريب جيد إلى حد ما معتبرين أن $f_1=\frac{1}{4}$.

(١٠-٦) المعاينة المثلى
وكسور المعاينة الجزئية

تعتمد هذه على نوع دالة التكلفة، وإذا كانت تكاليف السفر بين الوحدات غير مهمة، فإن إحدى الصيغ التي ثبتت فائدتها هي:

$$C=c_1n+c_2nm$$

والمركبة الأولى للتكلفة، c_1n ، متناسبة مع عدد الوحدات الأولية في العيّنة؛ والثانية، c_2nm ، متناسبة مع العدد الكلي لوحدات المرحلة الثانية أو العناصر. ومن النظرية (١٠-١) يمكن كتابة $V(\bar{y})$ على الشكل،

$$V(\bar{y})=\frac{1}{n}\left(S_1{}^2-\frac{S_2{}^2}{M}\right)+\frac{1}{mn}S_2{}^2-\frac{1}{N}S_1{}^2 \qquad (10.25)$$

ولا يعتمد الحد الأخير على اليمين على اختيار n و m . وجعل V أصغر ما يمكن مع C مثبتة، أو C أصغر ما يمكن مع V مثبت، يكافىء جعل الجداء التالي أصغر ما يمكن،

$$\left(V+\frac{1}{N}S_1{}^2\right)C=\left[\left(S_1{}^2-\frac{S_2{}^2}{M}\right)+\frac{S_2{}^2}{m}\right](c_1+c_2m)$$

ومن متراجحة كوشي - شوارتز نجد،

$$m_{opt}=\frac{S_2}{\sqrt{S_1{}^2-S_2{}^2/M}}\sqrt{c_1/c_2} \qquad (10.26)$$

شريطة أن يكون $S_1{}^2>S_2{}^2/M$. وندوّر m_{opt} إلى أقرب عدد صحيح ، وإذا كان m عددًا صحيحًا بحيث إن $m<m_{opt}<m+1$ فنقول، كقاعدة أفضل بقليل في حالة m_{opt} صغير (Cameron, 1951) دوّر إلى أعلى إذا كان $m_{opt}^2>m(m+1)$ ، وفيما عدا ذلك دوّر إلى الأسفل. وإذا كان $m_{opt}>M$ أو إذا كان $S_1{}^2<S_2{}^2/M$ خذ $m=M$ مستخدمًا معاينة بمرحلة واحدة. (الجداء $(V+S_1{}^2/N)C$ دالة متناقصة باطراد مع m عندما يكون $S_1{}^2<S_2{}^2/M$).

وفي معظم الحالات العملية يكون الحل الأمثل منبسطًا نسبيًا. وارتكاب خطأ بسيط في اختيار عدد الوحدات m لا يؤدي إلا إلى خسارة بسيطة في الدقة، كما يوضح هذا المثال التالي. ونكتب،

$$S_u{}^2=S_1{}^2-\frac{S_2{}^2}{M} \qquad (10.27)$$

مثـــال

ليكن،
$$c_1=10c_2, \qquad S_2=1.3S_u$$

فعندئذ،
$$m_{opt}=1.3\sqrt{10}=4.1$$

وسنعتبر التكلفة الإجمالية مثبتة، ثم نرى كيفية تغير تباين $\bar{y}$ مع m . ونفترض أن N كبير ومن (10.14) نكتب،

$$V(\bar{y})=\frac{S_u{}^2}{n}+\frac{S_2{}^2}{nm}$$
$$=\left(S_u{}^2+\frac{S_2{}^2}{m}\right)\frac{c_1+mc_2}{C}$$

حاذفين n بالاستفادة من معادلة التكلفة. وهذا يعطي ،

$$V(\bar{y})=\frac{S_u{}^2c_2}{C}\left(1+\frac{S_2{}^2}{mS_u{}^2}\right)\left(\frac{c_1}{c_2}+m\right)=\frac{S_u{}^2c_2}{C}\left(1+\frac{1.69}{m}\right)(10+m)$$

وبحذف العامل الثابت، يمكن حساب التباين النسبي من أجل قيم مختلفة لِـ m . ويبيّن الجدول (١٠-١) هذه التباينات والدقة النسبية (آخذين الدقة العظمى في حالة $m=4$ كمعيار).

ومع أي قيمة لِـ m بين 2 و 9 نجد أن الخسارة في الدقة بالنسبة إلى الدقة المثلى أقل من 12% .

جدول (١٠-١) التباينات النسبية والدقة النسبة من أجل قيم مختلفة لِـ m

$m =$	1	2	3	4	5	6	7	8	9	10
تباين نسبي	29.59	22.14	20.32	19.92	20.07	20.51	21.10	21.80	22.56	23.38
دقة نسبية	0.67	0.90	0.98	1.00	0.99	0.97	0.94	0.91	0.88	0.85

وفي الممارسة العملية، يستدعي اختيار m تقديرات لِـ c_1/c_2 ولِـ S_2/S_1 أو بصورة مكافئة لِـ S_2/S_u . وبسبب انبساط الحل الأمثل فلا حاجة للحصول على هذه النسب بدقة عالية. وإذا كانت c_1/c_2 معروفة بصورة حسنة إلى حد ما، وقمنا باختيار قيمة لِـ m ، ولنقل m_0 ، فهناك جدول مفيد (1955, Brooks) يقدم مدى لقيم S_2^2/S_u^2 ، وضمن حدود هذا المدى تعطي m_0 دقة لا تقل عن 90% من الدقة المثلى.

وتمّ الحصول على الجدول كما يلي. من أجل تكلفة معطاة، ومفترضين N كبيراً، وُجِد أن الدقة النسبية لِـ m_0 إلى m_{opt} هي،

$$\frac{V(\bar{y}|m_{opt})}{V(\bar{y}|m_0)} = \frac{(S_u\sqrt{c_1}+S_2\sqrt{c_2})^2}{S_u^2c_1+S_2^2c_2+m_0c_2S_u^2+c_1S_2^2/m_0} \qquad (10.28)$$

ومجموعة قيم S_2/S_u التي تتجاوز هذه العبارة من أجلها مستوى محددًا ما L ، هي تلك التي تقع بين الجذرين

$$\frac{S_2}{S_u} = \frac{\gamma \pm \sqrt{L(1-L)}(\sqrt{m_0}+\gamma^2/\sqrt{m_0})}{(L\gamma^2/m_0)-(1-L)} \qquad (10.29)$$

حيث $\gamma^2 = c_1/c_2$

والجدول (١٠-٢)، المقتبس بعد تكييفه من Brooks (1955) يبيّن الحدين الأعلى والأدنى لِـ S_2^2/S_u^2 في حالة $L=0.9$. والمجال الواسع بين الحدين الأعلى والأدنى

يلفت النظر في جميع الحالات تقريبًا. ونلاحظ أن مدى m_0 يتغير في أجزاء مختلفة من الجدول.

وإذا كان لدينا فكرة تقريبية عن قيم S_2^2/S_u^2 من أجل المفردات الرئيسة في مسح إحصائي، فيمكن استخدام الجدول (١٠-٢) لاختيار قيمة m_0. ونلاحظ أنه إذا كان ρ الارتباط بين عناصر من الوحدة نفسها، كما عرفناه في الفقرة (٩-٤)، فتكون النسبة S_2^2/S_u^2 مساوية تقريبًا لِـ $(1-\rho)/\rho$ وقيمة لِـ S_2^2/S_u^2 منخفضة إلى حد الواحد تقابل $\rho=0.5$. وقد يشكل هذا درجة عالية بصورة غير عادية للارتباط ضمن الوحدة. وبصورة مماثلة $\rho=0.1$ يعطي $S_2^2/S_u^2=9$، بينما يعطي $\rho=0.01$، $S_2^2/S_u^2=99$.

مثال

لنفرض أن c_1/c_2 حوالي الواحد وأننا نتوقع وقوع S_2^2/S_u^2 بين 5 و 100 من أجل المفردات الرئيسة. وتعطي الأعمدة $c_1/c_2=1$، القيمة $m_0=4$ كاختيار مُرضٍ باعتباره يغطي النسب من 4 إلى أكثر من 100 (في الواقع إلى 196). ومع $c_1/c_2=16$ والمدى المرغوب نفسه، يقترح الجدول قيمة لِـ m_0 في مكان ما بين 15 و 20. وبالإضافة إلى ذلك تبيّن الحسابات من (10.29) أن $m_0=18$ هي الأفضل. وهذا يغطي المدى من 5.2 إلى 84 (ليس متسعًا جدًّا كما هو مرغوب).

وعندما تكون تكلفة السفر بين الوحدات الأولية كبيرة، فقد تكون دالة التكلفة التالية أكثر دقة

$$C=c_1n+c_t\sqrt{n}+c_2nm \qquad (10.30)$$

باعتبار أن تكاليف السفر تميل إلى التناسب مع $\sqrt{n}$. وإذا حددنا قيمة مرغوبة لِـ $V(\bar{\bar{y}})$ فيمكن بسهولة أن نحسب من (10.25) الخاصة بالتباين $V(\bar{\bar{y}})$ الأزواج من القيم (n, m) التي تعطي ذلك التباين. وعندئذ تُحسب من (10.30) التكاليف من أجل تراكيب مختلفة، ونعثر بذلك على التركيب الذي يعطى أصغر تكلفة وعند تثبيت التكلفة سلفًا، يعطي Madow, Hurwitz, Hansen (1953) طريقة لتحديد التركيب

جدول (١٠-٢) مدى قيم S_2^2/S_u^2 التي تعطي m_0 في حدودها 90% على الأقل من الدقة العظمى

$c_1/c_2 =$	$\frac{1}{2}$		1		$c_1/c_2 =$	2		4	
m_0	L	U	L	U	m_0	L	U	L	U
1	0.0	11	0.0	4	2	0.5	8	0.2	4
2	2.0	98	1.1	22	3	1.2	21	0.5	8
3	4.1	>*	2.4	72	4	2.2	44	1.0	16
4	6.6	>	4.0	>	5	3.3	82	1.6	27
5	9.5	>	5.9	>	6	4.7	>	2.4	42
6	13	>	8.1	>	7	6.3	>	3.3	61
7	16	>	11	>	8	8.0	>	4.3	87
8	20	>	13	>	9	10	>	5.4	>
$c_1/c_2 =$	8		16		$c_1/c_2 =$	32		64	
m_0	L	U	L	U	m_0	L	U	L	U
6	1.0	17	0.3	8	5	0.1	3	0.0	2
7	1.5	24	0.5	11	10	0.4	12	0.1	7
8	2.0	32	0.7	15	15	1.2	26	0.3	14
9	2.6	42	1.0	19	20	2.7	46	0.7	24
10	3.3	53	1.3	23	25	4.5	74	1.5	37
15	7.6	>	3.5	55	30	6.9	>	2.5	52
20	13.3	>	6.6	>	35	9.7	>	3.7	71
25	20.4	>	10.5	>	40	13	>	5.2	93

* ترمز لـ "> 100."

(n, m) الذي يجعل التباين أصغر ما يمكن، كما يعطون جدولاً يسهّل الاختيار السريع. وتجدر ملاحظة أن n لديهم هي m عندنا والعكس بالعكس.

(١٠-٧) تقدير m_{opt} من مسح استطلاعي

نحصل أحياناً على تقديرين لـ S_2^2 و S_1^2 من مسح استطلاعي نختار فيه n' من الوحدات الأولية، ثم نأخذ m' عنصراً من كل وحدة. وتعالج هذه الفقرة اختيار n' و m' وإذا كان s_1^2 التباين بين متوسطات الوحدات و s_2^2 التباين بين الوحدات الجزئية ضمن الوحدات، كما عرفناهما في الفقرة (١٠-٤)، فإن (10.22) تعطي

$$E(s_1^2) = \left(S_1^2 - \frac{S_2^2}{M}\right) + \frac{S_2^2}{m'} = S_u^2 + \frac{S_2^2}{m'} \qquad (10.31)$$

ومع دالة تكلف بسيطة $c_1n + c_2nm$ نجد،

$$m_{opt} = \frac{S_2}{\sqrt{S_1^2 - S_2^2/M}}\sqrt{c_1/c_2}$$

وكتقدير لِـ m_{opt} من المسح الاستطلاعي، تقترح (10.31) أن نأخذ،

$$\hat{m}_{opt} = \frac{s_2}{\sqrt{s_1^2 - s_2^2/m'}}\sqrt{c_1/c_2} = \frac{\sqrt{m'}}{\sqrt{(m's_1^2/s_2^2) - 1}}\sqrt{c_1/c_2} \qquad (10.32)$$

والتقدير $\hat{m}_{opt}$ يخضع لخطأ معاينة يعتمد على خطأ المعاينة للنسبة s_1^2/s_2^2 ومن تحليل التباين نعلم أن $m's_1^2/s_2^2$ يتوزع وفق التوزيع،

$$F\left(1 + m'\frac{S_u^2}{S_2^2}\right)$$

حيث يمتلك F,$(n'-1)$ و $n'(m'-1)$ من درجات الحرية، شريطة أن يكون توزيع الـ y_{ij} هو التوزيع الطبيعي. وتقود هذه النتيجة إلى توزيع المعاينة لِـ $\hat{m}_{opt}$ من أجل قيم معطاة لِـ n' و m' أي أن،

$$\hat{m}_{opt} = \frac{\sqrt{m'c_1/c_2}}{\sqrt{F\left(1 + \frac{m'S_u^2}{S_2^2}\right) - 1}} \qquad (10.33)$$

مثال

في مثال الفقرة (١٠-٦) حيث،

$$c_1 = 10c_2, \qquad S_2 = 1.3S_u, \qquad m_{opt} = 1.3\sqrt{10} = 4.1$$

لندرس مدى جودة تقدير m_{opt} من عيّنة استطلاعية فيها $n'=10$ و $m'=4$ فمن (10.33) نجد،

$$\hat{m}_{opt} = \frac{6.324}{\sqrt{F[1 + (4/1.69)] - 1}} = \frac{6.324}{\sqrt{3.367F - 1}}$$

حيث يمتلك F ، 9 و30 درجـة من الحـرية . ولإيجاد حدود الفترة التي تغطّي قيمة $\hat{m}_{opt}$ في 80% من المرات ، نلاحظ من مستويي الأهمية وحيدي الذيل لِـ F بحجم 10% أن :

$$F_{.10}(9, 30) = 1.8490, \qquad F_{.90}(9, 30) = 1/F_{.10}(30, 9) = 1/2.2547 = 0.4435$$

وتعويض هذه القيم لِـ F يعطي ،

الحد الأعلى $\hat{m}_{opt}=9.0$ ، والحد الأدنى $\hat{m}_{opt}=2.8$ ،

وكما مبيّن في الجدول (١٠-١) فإن أيّ m في هذا المدى يعطي دقة قريبة من المثلى . وهكذا، ومع $n'=10$ ، $m'=4$ كانت فرص نقص بسيط في الدقة، في بيان إحصائي طبيعي ، هي ثمانية من عشرة .

وتبدو حدود الـ 80 و 95% في حالة $n'=5, 10, 20$ و $m'=4$ في الجدول (١٠-٣) . ومع $n'=20$ ، نكون متأكدين تقريبًا من تقدير m_{opt} بدقة قريبة من المثلى . وليس الأمر كذلك في حالة $n'=5$.

جدول (١٠-٣) الحدود العليا والدنيا من أجل $\hat{m}_{opt}$

n'	80%	90%
5	2.5, ∞	1.8, ∞
10	2.8, 9.0	2.3, ∞
20	3.1, 6.4	2.7, 9.1

وإذا كانت النسبـة c_1/c_2 نفسها في المسح الاستطلاعي كما في المسح الرئيس، فستكـون تكلفـة المسـح الاستـطلاعي متنـاسبـة مع $c_1n'+c_2n'm'$ ويعـطي Brooks (1955) جدول القيم (n',m') في المسح الاستطلاعي الأكثر وفرًا الذي يقدم دقة نسبية متوقعة في تقدير m_{opt} تبلغ 90% . ويبيّن الجدول (١٠-٤) جزءًا من هذا الجدول .

وتفترض الحسابات أن N و M كبيران : وتكون التصميات محافظة إذا أخذنا في الاعتبار حدود الـ ت م م . ونلاحظ أننا لا نحتاج إلى أكثر من 10 وحدات أولية وأن التصاميم غير حسّاسة إلى حد ما بالنسبة لقيم النسبة c_1/c_2 .

جدول (١٠-٤) تصميمات عيّنة استطلاعية تمتلك دقة نسبية متوقعة تبلغ 90% .

c_1/c_2	≤1		2		4		8		16		32		64	
S_2^2/S_u^2	n'	m'	n'	m'	n'	m'	n'	m'	n'	m'	n'	m'	n'	m'
1	7	3	6	4	6	5	5	6	5	7	4	10	4	12
2	8	5	7	7	6	9	6	9	5	13	5	14	4	20
4	9	9	8	11	8	12	7	14	7	15	5	25	5	27
8	10	14	10	15	9	17	9	18	8	22	6	32	5	44
16	10	25	10	27	10	27	10	28	8	37	7	46	6	60
32	10	46	10	47	10	48	10	49	9	58	8	69	6	102
64	10	92	10	93	10	96	10	100	10	104	8	137	7	169

(١٠ ـ ٨) معاينة على ثلاث مراحل

ننقـل أحيانـًا عملية المعاينة الجزئية إلى مرحلة ثالثة بأخذ عيّنة من الوحدات الجزئية بدلاً من تعدادها بالكامل. وعلى سبيل المثال، في مسوح إحصائية لتقدير إنتاج محصول معيّن في الهند (Sukhatme, 1947) شكلت القرية وحدة معاينة مريحة. وضمن القرية، اقتُصر على اختيار بعض الحقول المزروعة بهذا المحصول، بحيث يصبح الحقل وحدة جزئية. وعند اختيار حقل تُجنى منه أجزاء معيّنة فقط لتحديد إنتاج الفدان؛ وهكذا فإننا نأخذ عيّنة من الوحدة الجزئية نفسها. وفي حال وجود تحليل فيزيائي أو كيميائي للمحصول فمن الممكن استخدام معاينة أخرى باعتبار أن مثل هذه التحاليل تتم غالبًا على جزء من العيّنة التي أخذناها من الحقل.

والنتـائـج هي تعميم مبـاشر لتلك المتعلقـة بمعـاينة على مرحلتين وسنعطيها باختصـار. ويتضمن المجتمـع N من وحـدات المرحلة الأولى، وفي كل منها M من وحدات المرحلة الثانية، وفي كل منها K من وحدات المرحلة الثالثة. والأعداد الموافقة للعيّنة هي k,m,n على الترتيب. لتكن y_{iju} القيمة التي نحصل عليها من الوحدة u من وحدات المرحلة الثالثة الموجودة في الوحدة j من وحدات المرحلة الثانية المسحوبة من الوحدة الأولية i . ومتوسطات المجتمع الموافقة على أساس وحدة المرحلة الثالثة هي كما يلي:

جدول (١٠-٤) تصميمات عيّنة استطلاعية تمتلك دقة نسبية متوقعة تبلغ 90% .

c_1/c_2	≤ 1		2		4		8		16		32		64	
S_2^2/S_u^2	n'	m'	n'	m'	n'	m'	n'	m'	n'	m'	n'	m'	n'	m'
1	7	3	6	4	6	5	5	6	5	7	4	10	4	12
2	8	5	7	7	6	9	6	9	5	13	5	14	4	20
4	9	9	8	11	8	12	7	14	7	15	5	25	5	27
8	10	14	10	15	9	17	9	18	8	22	6	32	5	44
16	10	25	10	27	10	27	10	28	8	37	7	46	6	60
32	10	46	10	47	10	48	10	49	9	58	8	69	6	102
64	10	92	10	93	10	96	10	100	10	104	8	137	7	169

(١٠ ـ ٨) معاينة على ثلاث مراحل

ننقـل أحيانـًا عملية المعاينة الجزئية إلى مرحلة ثالثة بأخذ عيّنة من الوحدات الجزئية بدلاً من تعدادها بالكامل. وعلى سبيل المثال، في مسوح إحصائية لتقدير إنتاج محصول معيّن في الهند (Sukhatme, 1947) شكلت القرية وحدة معاينة مريحة. وضمن القرية، اقتُصر على اختيار بعض الحقول المزروعة بهذا المحصول، بحيث يصبح الحقل وحدة جزئية. وعند اختيار حقل تُجنى منه أجزاء معيّنة فقط لتحديد إنتاج الفدان؛ وهكذا فإننا نأخذ عيّنة من الوحدة الجزئية نفسها. وفي حال وجود تحليل فيزيائي أو كيميائي للمحصول فمن الممكن استخدام معاينة أخرى باعتبار أن مثل هذه التحاليل تتم غالبًا على جزء من العيّنة التي أخذناها من الحقل.

والنتـائـج هي تعميم مبـاشر لتلك المتعلقـة بمعـاينة على مرحلتين وسنعطيها باختصـار. ويتضمن المجتمـع N من وحـدات المرحلة الأولى، وفي كل منها M من وحدات المرحلة الثانية، وفي كل منها K من وحدات المرحلة الثالثة. والأعداد الموافقة للعيّنة هي k,m,n على الترتيب. لتكن y_{iju} القيمة التي نحصل عليها من الوحدة u من وحدات المرحلة الثالثة الموجودة في الوحدة j من وحدات المرحلة الثانية المسحوبة من الوحدة الأولية i . ومتوسطات المجتمع الموافقة على أساس وحدة المرحلة الثالثة هي كما يلي:

$$\bar{Y}_{ij} = \frac{\sum_{u}^{K} y_{iju}}{K}, \quad \bar{Y}_i = \frac{\sum_{j}^{M}\sum_{u}^{K} y_{iju}}{MK}, \quad \bar{\bar{Y}} = \frac{\sum_{i}^{N}\sum_{j}^{M}\sum_{u}^{K} y_{iju}}{NMK}$$

والمطلوب تباينات المجتمع التالية.

$$S_1{}^2 = \frac{\sum_{i}^{N} (\bar{Y}_i - \bar{\bar{Y}})^2}{N-1}$$

$$S_2{}^2 = \frac{\sum_{i}^{N}\sum_{j}^{M} (\bar{Y}_{ij} - \bar{Y}_i)^2}{N(M-1)}$$

$$S_3{}^2 = \frac{\sum_{i}^{N}\sum_{j}^{M}\sum_{u}^{K} (y_{ijk} - \bar{Y}_{ij})^2}{NM(K-1)}$$

نظرية (١٠-٣)

إذا استُخدمت المعاينة العشوائية البسيطة في المراحل الثلاث جميعها، يكون متوسط العيّنة $\bar{\bar{y}}$ لكل من وحدات المرحلة الثالثة تقديرًا غير منحاز لِـ $\bar{\bar{Y}}$ بتباين يساوي،

$$V(\bar{\bar{y}}) = \frac{1-f_1}{n} S_1{}^2 + \frac{1-f_2}{nm} S_2{}^2 + \frac{1-f_3}{nmk} S_3{}^2 \qquad (10.34)$$

حيث $f_1 = n/N, f_2 = m/M, f_3 = k/K$ هي كسور المعاينة في المراحل الثلاث.

برهان

نشير فقط إلى الخطوات الرئيسة. فلنكتب،

$$\bar{\bar{y}} - \bar{\bar{Y}} = (\bar{\bar{y}} - \bar{\bar{Y}}_{nm}) + (\bar{\bar{Y}}_{nm} - \bar{\bar{Y}}_n) + (\bar{\bar{Y}}_n - \bar{\bar{Y}}) \qquad (10.35)$$

حيث $\bar{\bar{Y}}_{nm}$ متوسط المجتمع لِـ nm من وحدات المرحلة الثانية التي اختيرت، و $\bar{\bar{Y}}_n$ متوسط المجتمع للوحدات الأولية الـ n التي اختيرت. وعندما نربع ونأخذ المتوسط، تنعدم الحدود الجدائية. وتتمخض مساهمات الحدود التربيعية عما يلي،

$$E(\bar{\bar{y}} - \bar{\bar{Y}}_{nm})^2 = \frac{1-f_3}{nmk} S_3^2$$

$$E(\bar{\bar{Y}}_{nm} - \bar{\bar{Y}}_n)^2 = \frac{1-f_2}{nm} S_2^2$$

$$E(\bar{\bar{Y}}_n - \bar{\bar{Y}})^2 = \frac{1-f_1}{n} S_1^2$$

وعند إضافة الحدود الثلاثة نحصل على المطلوب.

نظرية (١٠-٤)

تقدير العيّنة غير المنحاز لـ $V(\bar{\bar{y}})$ هو،

$$v(\bar{\bar{y}}) = \frac{1-f_1}{n} s_1^2 + \frac{f_1(1-f_2)}{nm} s_2^2 + \frac{f_1 f_2(1-f_3)}{nmk} s_3^2 \qquad (10.36)$$

حيث s_1^2, s_2^2, s_3^2 مقادير العيّنة المقابلة لـ S_1^2, S_2^2, S_3^2 على الترتيب.

برهان

يمكن برهان هذا باستخدام الطرق المذكورة في الفقرة (١٠-٤) أو بصورة بديلة ببرهان أن

$$E(s_1^2) = S_1^2 + \frac{1-f_2}{m} S_2^2 + \frac{1-f_3}{mk} S_3^2 \qquad (10.37)$$

$$E(s_2^2) = S_2^2 + \frac{1-f_3}{k} S_3^2$$

و $E(s_3^2) = S_3^2$. وللحصول على النتيجة الأولى، لنرمز بـ $\bar{y}_{iK}$ للمتوسط فوق وحدات المرحلة الثانية في الوحدة الأولية i ، علمًا أن جميع العناصر الـ K قد أحصيت في المرحلة الثالثة. وليكن $\bar{\bar{y}}_K$ متوسط القيم $\bar{y}_{iK}$ الـ n . وعندئذ، ومن (10.22) في معاينة على مرحلتين، نستنتج أن،

$$E\left[\frac{\sum^n (\bar{y}_{iK} - \bar{\bar{y}}_K)^2}{n-1}\right] = S_1^2 = \frac{1-f_2}{m} S_2^2 \qquad (10.38)$$

والآن، إذا كان $\bar{\bar{y}}_i$ متوسط العيّنة من الوحدة الأولية i ، نكتب،

$$(\bar{\bar{y}}_i - \bar{\bar{\bar{y}}}) = (\bar{\bar{y}}_{iK} - \bar{\bar{\bar{y}}}_K) + [(\bar{\bar{y}}_i - \bar{\bar{y}}_{iK}) - (\bar{\bar{\bar{y}}} - \bar{\bar{\bar{y}}}_K)] \qquad (10.39)$$

وبأخذ المتوسطات أولاً فوق عيّنات ثبّتنا فيها وحدات المرحلة الأولى والمرحلة الثانية، يمكن البرهان على أن،

$$\frac{1}{(n-1)} E \sum^{n} [(\bar{\bar{y}}_i - \bar{\bar{y}}_{iK}) - (\bar{\bar{\bar{y}}} - \bar{\bar{\bar{y}}}_K)]^2 = \frac{(1-f_3)S_3^{\,2}}{mk} \qquad (10.40)$$

ولا تُسهم الحدود الجدائية من (10.39) بشيء. وهذا يثبت النتيجة الخاصة بِـ $E(s_1^{\,2})$. ونجد بصورة مماثلة النتيجة الموافقة لِـ $E(s_2^{\,2})$. وبالتالي،

$$\begin{aligned} E[v(\bar{\bar{\bar{y}}})] &= \frac{1-f_1}{n}\left(S_1^{\,2} + \frac{1-f_2}{m} S_2^{\,2} + \frac{1-f_3}{mk} S_3^{\,2}\right) \\ &\quad + \frac{f_1(1-f_2)}{nm}\left(S_2^{\,2} + \frac{1-f_3}{k} S_3^{\,2}\right) + \frac{f_1 f_2 (1-f_3)}{nmk} S_3^{\,2} \\ &= \frac{1-f_1}{n} S_1^{\,2} + \frac{1-f_2}{nm} S_2^{\,2} + \frac{1-f_3}{nmk} S_3^{\,2} = V(y) \qquad (10.41) \end{aligned}$$

وكما في المعاينة على مرحلتين، يتضح من (10.36) أنه إذا كان f_1 مهملاً فإن $v(\bar{\bar{\bar{y}}})$ يُختزل إلى،

$$v(\bar{\bar{\bar{y}}}) = \frac{s_1^{\,2}}{n} = \frac{\sum^{n} (\bar{\bar{y}}_i - \bar{\bar{\bar{y}}})^2}{n(n-1)} \qquad (10.42)$$

وهذا التقدير محافظ إذا لم يكن f_1 مهملاً.

ومع دالة تكلفة صيغتها:

$$C = c_1 n + c_2 nm + c_3 nmk \qquad (10.43)$$

تكون القيم المثلى لِـ k و m هي،

$$k_{opt} = \frac{S_3}{\sqrt{S_2^{\,2} - S_3^{\,2}/K}} \sqrt{c_2/c_3}, \qquad m_{opt} = \frac{\sqrt{S_2^{\,2} - S_3^{\,2}/K}}{\sqrt{S_1^{\,2} - S_2^{\,2}/M}} \sqrt{c_1/c_2} \qquad (10.44)$$

وينبغي أن يكون تعميم النتائج في هذه الفقرة إلى مراحل معاينة إضافية واضحًا وذلك من بنية الصيغ المعطاة.

(١٠-٩) معاينة طبقية للوحدات

يمكن أن تترافق المعاينة الجزئية بأي نوع من المعاينة للوحدات الأولية. ويمكن للمعاينة الجزئية نفسها أن تستخدم التقسيم إلى طبقات أو المعاينة النمطية. ويمكن بناء علاقات التباين الموافقة لهذه التعديلات من العلاقات الموافقة للطرق الأبسط.

والنتائج معطاة في حالة معاينة طبقية للوحدات الأولية في عيّنة على مرحلتين. ونفترض أن حجوم الوحدات الأولية ثابتة ضمن طبقة معطاة، ولكنها يمكن أن تختلف من طبقة إلى طبقة. وتحصل هذه الحالة عند تقسيم الوحدات الأولية إلى طبقات وفقًا للحجم وهكذا تصبح الحجوم ضمن طبقة، ثابتة أو ثابتة تقريبًا.

وتتضمن الطبقة الـ h عددًا من الوحدات الأولية $= N_h$ وفي كل منها M_h من وحدات المرحلة الثانية؛ وأعداد العيّنة المقابلة n_h و m_h. وتقدير متوسط المجتمع لكل وحدة من وحدات المرحلة الثانية هو،

$$\bar{\bar{y}}_{st} = \frac{\sum_h N_h M_h \bar{\bar{y}}_h}{\sum_h N_h M_h} = \sum_h W_h \bar{\bar{y}}_h \qquad (10.45)$$

حيث $W_h = N_h M_h / \Sigma N_h M_h$ هو الحجم النسبي للطبقة بدلالة وحدات المرحلة الثانية، و $\bar{\bar{y}}_h$ هو متوسط العيّنة في الطبقة. وبتطبيق النظرية (١٠-١) ضمن كل طبقة، نجد،

$$V(\bar{\bar{y}}_{st}) = \sum_h W_h^2 \left(\frac{1-f_{1h}}{n_h} S_{1h}^2 + \frac{1-f_{2h}}{n_h m_n} S_{2h}^2 \right) \qquad (10.46)$$

حيث $f_{1h} = n_h/N_h, f_{2h} = m_h/M_h$

ومن النظرية (١٠-٢) نجد تقدير العيّنة غير المنحاز

$$v(\bar{\bar{y}}_{st}) = \sum_h W_h^2 \left[\frac{1-f_{1h}}{n_h} s_{1h}^2 + \frac{f_{1h}(1-f_{2h})}{n_h m_h} s_{2h}^2 \right] \qquad (10.47)$$

ونحصل على التباينات الموافقة لتقدير مجموع المجتمع بضرب كل من العلاقتين (10.46) و (10.47) بـ $(\Sigma N_h M_h)^2$.

(١٠-١٠) محاصّة مثلى في حالة معاينة طبقية

نعـالج هنا أفضل اختيار للمقادير n_h و m_h . وإذا لم تكن تكاليف السفر بين الوحدات عاملاً رئيسًا، فيمكن تمثيل التكلفة بصورة مناسبة بالعلاقة:

$$C = \sum_h c_{1h} n_h + \sum_h c_{2h} n_h m_h \tag{10.48}$$

ومن (10.46) يمكن كتابة التباين على الشكل،

$$V(\bar{\bar{y}}_{st}) = \sum_h W_h^2 \left[\frac{1}{n_h} \left(S_{1h}^2 - \frac{S_{2h}^2}{M_h} \right) + \frac{1}{n_h m_h} S_{2h}^2 - \frac{1}{N_h} S_{1h}^2 \right]$$

والكمية،

$$V(\bar{\bar{y}}_{st}) + \lambda \left(\sum_h c_{1h} n_h + \sum_h c_{2h} n_h m_h - C \right)$$

حيث λ مضروب لاغـرانـج، هي دالّـة في المقادير n_h و $(n_h m_h)$. وبالتالي نجد عند جعل V أصغر ما يمكن مع C مثبتة، أو العكس بالعكس،

$$n_h \sqrt{\lambda} = \frac{W_h}{\sqrt{c_{1h}}} \sqrt{S_{1h}^2 - S_{2h}^2 / M_h} \tag{10.49}$$

$$n_h m_h \sqrt{\lambda} = \frac{W_h S_{2h}}{\sqrt{c_{2h}}} \tag{10.50}$$

وهذه تعطي،

$$m_h = \frac{S_{2h}}{\sqrt{S_{1h}^2 - S_{2h}^2 / M_h}} \sqrt{c_{1h} / c_{2h}} \tag{10.51}$$

والعلاقة الموافقة لِـ m_h المثلى هي بالضبط العلاقة نفسها التي نجدها في معاينة غير طبقية [(10.26)6 من الفقرة (١٠-٦)].

ومن (10.49) ، وباعتبار $W_h \propto N_h M_h$ نجد،

$$n_h \propto \frac{N_h M_h S_{uh}}{\sqrt{c_{1h}}} \quad \text{حيث} \quad S_{uh}^2 = S_{1h}^2 - \frac{S_{2h}^2}{M_h} \tag{10.52}$$

وبـما أن التقـديرات ذاتية الـترجيح مريحة، فسنتساءل عن الظروف التي تقود

المحاصّة المثلى تحتها إلى تقدير ذاتي الترجيح. ومن (10.45) نستنتج أن $\bar{y}_{st}$ ذاتية الترجيح إذا كان ثابت $= n_h m_h / N_h M_h = f_0$، باعتبار أنه في هذه الحالة،

$$\bar{\bar{y}}_{st} = \frac{\sum_h (N_h M_h / n_h m_h) \sum_i^{n_h} \sum_j^{m_h} y_{hij}}{\sum N_h M_h} = \frac{\sum_h \sum_i \sum_j y_{hij}}{f_0 \sum N_h M_h}$$

$$= \frac{\sum_h \sum_i \sum_j y_{hij}}{\sum n_h m_h} = \bar{\bar{y}}$$

والشرط، كما قد نتوقع، هو بقاء f_0، أي بقاء احتمال اختيار وحدة جزئية، ثابتًا في جميع الطبقات.

ومن (10.50) تعطي المحاصة المثلى،

$$f_{0h} = \frac{n_h m_h}{N_h M_h} \propto \frac{S_{2h}}{\sqrt{c_{2h}}} \qquad (10.53)$$

وكثيرًا ما تكون c_{2h}، التكلفة لكل وحدة من وحدات المرحلة الثانية، هي نفسها تقريبًا في وحدات أولية كبيرة وصغيرة؛ ولكن يمكن أن يكون S_{2h} أكبر في الوحدات الكبيرة منه في الوحدات الصغيرة. وعلى أي حال، وباعتبار أن الوضع الأمثل غير حساس، فغالبًا ما ستكون عيّنة ذاتية الترجيح في دقة العيّنة المثلى نفسها تقريبًا. ونلاحظ أن هذه النتيجة تصحّ حتى إذا كانت المعاينة المثلى للوحدات الأولية بعيدة عن واقع التناسب.

تمـاريــن

(١٠-١) خُزنت مجموعة من 20,000 سجل في 400 من جرارات الأضابير، كل منها يتضمن 50 سجلاً. وفي معاينة على مرحلتين، سُحبت خمسة سجلات عشوائيًا من كل من 80 جرارًا اختيرت عشوائيًا. ومن أجل مفردة واحدة، كانت تقديرات التباين $s_1^2 = 362, s_2^2 = 805$ كما عرفناها في الفقرة (١٠-٤). (ا) احسب الخطأ المعياري للمتوسط لكل سجل من هذه العيّنة. (ب) قارن هذا مع الخطأ المعياري المعطى بالعلاقة التقريبية (10.23) في الفقرة (١٠-٤).

(١٠ـ٢) من نتـائـج عيّنـة استـطلاعية على مرحلتـين، اختيرت فيها m' من الوحدات الجزئية من كل من n' وحدة أولية، ومن المفيد أن نستطيع تقدير قيمة $V(\bar{y})$ التي ستعطيها عيّنة لاحقة تتضمن m من الوحدات الجزئية من كل من n من الوحدات الأولية. بين أن،

$$\hat{V}(\bar{y}) = \left(\frac{N-n}{N}\right)\frac{s_1^2}{n} + \frac{s_2^2}{mn}\left(1 - \frac{m}{m'} + \frac{mn}{m'N} - \frac{mn}{MN}\right)$$

تقـديـر غير منحــاز لِـ $V(\bar{y})$ حيث حُسبـت s_2^2 و s_1^2 من الـعيّنــة التمهيدية.

تلميح: استخدم النظرية (١٠ـ١) والنتيجة (10.22).

$$E(s_1^2) = S_1^2 - \frac{S_2^2}{M} + \frac{S_2^2}{m}$$

(١٠ـ٣) في معـاينـة حقـول القمـح في كَنساس، حيث الحقل وحدة أولية، يقدّم King و McCarty (1941) متوسطات المربعات التالية للإنتاج بالبوشل لكل فدان: $s_1^2 = 165, s_2^2 = 66$ وأخــذت عيّنتـان جزئيتـان من كل حقـلٍ، قارن في حالـة عيّنـة من n حقلاً، تباينات متوسط العيّنة كما تعطيها (ا) العيّنة كما أُخذت فعلاً، (ب) أربع عيّنات جزئية من كل حقل من الحقول الـ n ، (ج) جني الحقول الـ n بالكامل.

يمكن أن نفترض N و M كبيرين وثابتين. وفي (ج) نفترض أن الجني الكامل مكافىء لمعاينة بمرحلة واحدة (أي للحالة $m=M$).

(١٠ـ٤) في المسح نفسه، ومع عيّنتين جزئيتين لكل حقل، كانت متوسطات المربعات للنسب المئوية للبروتين $s_1^2 = 7.73$ ، $s_2^2 = 1.43$. ما هو عدد الحقول المطلوبة لتقـدير متـوسط الإنتـاج في حدود 1± بوشــل، ومتـوسط النسبـة المئوية للبروتين في حدود $\frac{1}{4}$± وذلـك باستثناء فرصة واحدة من عشرين في كل حالة؟ أنجز الحسابات (ا) مفـترضًا أخذ عيّنتين جزئيتين لكل حقل في المسح الرئيس، (ب) مفترضًا الجني الكامل لحقل في المسح الرئيس.

(١٠ـ٥) في بيان إنتاج القمح في التمرين (١٠ـ٣)، ما هي قيمة c_1/c_2 في دالة تكلفة خطية، إذا كان تقدير القيمة المثلى لِـ m هو 2 ؟

(١٠ـ٦) إذا كان كل من m/M و n/N صغيراً، وكانت دالة التكلفة خطية، فبيّن أن $m=2$ يعطي قيمة للتباين $V(\bar{y})$ أصغر من تلك التي يعطيها $m=1$ إذا كان،

$$\frac{c_1}{c_2} > 2\frac{S_1^2}{S_2^2}$$

(١٠ـ٧) يتعامل متجر كبير مع 20,000 حساب تُدفع كل شهر، ويجري التحقق من عيّنة 2% ($m=400$) كل شهر فوق فترة سنتين ($n=24$). وقد وُجد أن عدد الحسابات التي تتضمن خطأ لكل شهر (من بين 400 حسابًا) هي (مرتبة وفقًا لقيمها من الأصغر إلى الأكبر) 0, 0, 1, 1, 2, 4, 4, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9, 9, 10, 10, 13, 14, 17، والترتيب وفق الزمن غير منتظم. احسب s_1^2 and s_2^2 من نتائج الفقرة (١٠ـ٥). وبالتالي احسب الخطأ المعياري لِـ $\bar{p}$، كتقدير للنسبة المئوية من الحسابات التي تتضمن خطأ فوق فترة سنة، والذي يمكن الحصول عليه من تحقيق (ا) 1200 حساب من شهر بمفرده، اخترناه عشوائيًا. (ب) 300 حساب من كل من 4 أشهر اخترناها عشوائيًا، (جـ) 100 حساب كل شهر. **تلميح**: استخدم إما العلاقة في التمرين (١٠ـ٢) مع $m'=400$ أو احصل على تقديرات غير منحازة لِـ S_2^2 و S_1^2 ثم استخدم النظرية (١٠ـ١).

(١٠ـ ٨) عند التخطيط للمسح على مرحلتين توقعنا أن تكون c_1/c_2 حوالي 4 وأن S_2^2/S_u^2 قد تقع بين 5 و 50. (ا) ما هي القيمة التي تختارها لِـ m من الجدول (١٠ـ٢)؟ (ب) لنفرض أنه بعد إتمام المسح وجدنا c_1/c_2 قريبة من 8، وأن S_2^2/S_u^2 كانت حوالي 25. احسب الدقة النسبية التي تعطيها m التي اخترتها بالمقارنة مع الدقة التي تعطيها m المثلى. (جـ) قم بالحسابات نفسها في حالة $S_2^2/S_u^2=100$، $c_1/c_2=4$.

(١٠ـ٩) إذا كان ρ معامل الارتباط بين وحدات المرحلة الثانية في الوحدة الأولية نفسها. فبرهن أن،

$$\frac{1-\rho}{\rho} = \frac{S_2^2}{[(N-1)/N]S_1^2 - S_2^2/M} \doteq \frac{S_2^2}{S_u^2}$$

[وهذا يثبت نتيجة استخدمت في الفقرة (١٠-٦)].

(١٠-١٠) بيّن أنه إذا كان $s_u^2>0$، وفق رموز الفقرة (١٠-٦)، فإن عيّنة عشوائية بسيطة من n وحدة أولية، مع اختيار عنصر واحد من كل وحدة، تكون أكثر دقة من عيّنة عشوائية بسيطة من n عنصرًا $(M>1, n>1)$. بيّن أن دقتي الطريقتين متساويتان إذا كان n/N مهملاً. هل تتوقع هذا بالبداهة؟

الفصل الحادي عشر

المعاينة الجزئية بوحدات غير متساوية الحجم

(١١-١) مقدمة

عند معاينة مجتمعات كبيرة، كثيراً ما نواجه وحدات أولية تتغير من حيث حجمها. وعلى أي حال، فغالبًا ما تُملي اعتبارات التكلفة استخدام المعاينة متعددة المراحل، وهكذا تكون المسائل التي نناقشها في هذا الفصل متواترة الوقوع. وإذا كانت الحجوم لا تتغير كثيراً، فإحدى الطرق هي أن نقسم إلى طبقات وفقًا لحجم الوحدة الأولية، بحيث تصبح الوحدات ضمن طبقة واحدة متساوية في حجومها، أو تقريبًا كذلك. وعندئذ يمكن أن تشكل العلاقات في الفقرة (١٠-٩) تقريبًا مناسبًا. وعلى أي حال، فغالبًا ما تبقى فروق كبيرة في الحجم ضمن بعض الطبقات، وأحيانًا يكون من المستحسن أن نبني التقسيم إلى طبقات على متغيرات أخرى. وفي مجلة دورية للمسوح الإحصائية الاجتماعية البريطانية، وهي عيّنات تغطي البلاد بأكملها معتبرة المناطق كوحدات أولية، أشار Gray و Corlett (1950) إلى أن الحجم قد اعتُبر في البداية أحد المتغيرات التي يُقسم بموجبها إلى طبقات، ولكن وُجد أن هناك عاملاً آخر أفضل من الحجم، وذلك عندما أصبحت الخواص المميزة للمجتمع معروفة بصورة أفضل.

ونحتاج إلى بعض الجهود المركزة لكي نحصل على معرفة فاعلة بالمعاينة متعددة المراحل، عندما تتغير الوحدات من حيث حجمها، وذلك بسبب مرونة الطريقة. ويمكن اختيار الوحدات، إما باحتمالات متساوية، أو باحتمالات متناسبة مع الحجم، أو مع تقدير ما للحجم. ويمكن استنباط القواعد المختلفة لتحديد المعاينة وكسور

المعـاينة الجزئية، كما تتوافر طرق مختلفة للتقدير. وتعتمد فوائد الطرق المختلفة على طبيعـة المجتمـع، وعـلى التكـاليف الميدانية، وعـلى المعلومات الإحصائية الإضافية الموجودة تحت تصرفنا.

والجـزء الأول من هذا الفصـل مكـرّس لوصف الطرق الرئيسة التي هي قيد الاستخدام. وسنبدأ بمجتمع يتألف من طبقة بمفردها. ويمكن التعميم إلى معاينة طبقية، كما في الفصول السابقة، وذلك بجمع علاقات التباين الملائمة فوق الطبقات. وللتبسيط نفترض أولاً أننا اخترنا فقط وحدة أولية بمفردها، أي $n=1$. وهذه الحالة ليست غير عملية بالمرة كما تبـدو للوهلة الأولى، لأنـه عنـدمـا يوجد عدد كبير من الطبقات، فقد نتمكن من إحراز دقة مُرضية في التقدير حتى ولو كان $n_h=1$. وفي المسوح الإحصائية الشهرية التي يأخذها مكتب الإحصاء في الولايات المتحدة، والتي تهدف إلى تقدير عدد العاملين، نجد أن الوحدة الأولية هي المنطقة أو زمرة من المناطق المتجاورة، وهذه الوحدة كبيرة، إلا أن لها ميزات إدارية تخفض من التكاليف. وبما أن المناطق بعيدة عن أن تكون منتظمة في خواصها الرئيسة، فقد جرى التوسّع في التقسيم إلى طبقات وذلك إلى الحد الذي اختيرت معه منطقة واحدة فقط من كل طبقة. وبالتالي فإنـه في خطة معـاينـة كهذه، تكون النظرية التي سنناقشها هنا قابلة للتطبيق على طبقة بمفردها.

وكـما في الفصـول السـابقة، فإن الكميات المطلوب تقديرها يمكن أن تكون مجموع المجتمع Y ، أو متوسط المجتمع (عادة المتوسط لكل وحدة جزئية $\bar{\bar{Y}}$)، أو نسبة متغيرين.

رمـوز: نرمز بـ y_{ij} للملاحظة من الوحدة الجزئية j ضمن الوحدة i. وتشير الرموز التالية إلى الوحدة i :

	مجتمع	عيّنة
عدد الوحدات الجزئية	M_i	m_i
المتوسط لكل وحدة جزئية	$\bar{Y}_i$	$\bar{y}_i$
المجموع	$Y_i = M_i\bar{Y}_i$	$y_i = m_i\bar{y}_i$

كما تشير الرموز التالية إلى المجتمع ككل أو إلى عيّنة:

	مجتمع	عيّنة
عدد الوحدات الجزئية	$M_0 = \sum^N M_i$	$\sum^n m_i$
المجموع	$Y = \sum^N Y_i$	$\sum^n y_i$
المتوسط لكل وحدة جزئية	$\bar{\bar{Y}} = Y/M_0$	$\bar{\bar{y}} = \sum y_i / \sum m_i$
المتوسط لكل وحدة أولية	$\bar{Y} = Y/N$	$\bar{y} = \sum y_i / n$

(١١-٢) طرق المعاينة عندما يكون n=1

لنفرض أننا اخترنا الوحدة i ، وأنها تحوي M_i من الوحدات التي أخذنا منها عيّنة عشوائية من m_i وحدة. فسنقدم الآن ثلاث طرق لتقدير $\bar{\bar{Y}}$ ، المتوسط لكل وحدة جزئية أو وحدة المرحلة الثانية كما تدعى غالبًا.

I ـ وحدات اختيرت باحتمالات متساوية

$$\text{تقدير} = \bar{y}_1 = \bar{y}_i.$$

والتقدير هو متوسط العيّنة لكل وحدة جزئية وهو تقدير منحاز. ذلك لأنه عند تكرار المعاينة من الوحدة نفسها، فإن متوسط $\bar{y}_i$ هو $\bar{Y}_i$ وبما أن لكل وحدة الفرصة نفسها في أن تكون الوحدة المختارة، فإن متوسط $\bar{Y}_i$ هو:

$$\frac{1}{N}\sum_{i=1}^{N} \bar{Y}_i = \bar{\bar{Y}}_a \qquad \text{(مثلاً)}$$

ولكن متوسط المجتمع هو:

$$M = \sum_{i=1}^{N} M_i \qquad \text{حيث} \quad \bar{\bar{Y}} = \frac{\sum_{i=1}^{N} M_i \bar{Y}_i}{M},$$

وبالتالي يكون الانحياز $(\bar{\bar{Y}}_a - \bar{\bar{Y}})$. وبما أن الطريقة منحازة فسنحسب متوسط مربعات الخطأ (MSE) حول $\bar{\bar{Y}}$. لنكتب،

$$\bar{y}_i - \bar{\bar{Y}} = (\bar{y}_i - \bar{Y}_i) + (\bar{Y}_i - \bar{\bar{Y}}_a) + (\bar{\bar{Y}}_a - \bar{\bar{Y}})$$

ثم لنربع ونأخذ التوقع فوق جميع العيّنات الممكنة. فكل مساهمات الحدود الجداثية هي الصفر وتوقعات الحدود المربعة تتبع بسهولة من الطرق المعطاة في الفصل العاشر.

ونجد:

$$\text{MSE}(\bar{y}_1) = \frac{1}{N}\sum_{i=1}^{N}\frac{(M_i - m_i)}{M_i}\frac{S_{2i}^2}{m_i} + \frac{1}{N}\sum_{i=1}^{N}(\bar{Y}_i - \bar{\bar{Y}}_a)^2 + (\bar{\bar{Y}}_a + \bar{\bar{Y}})^2 \quad (11.1)$$

حيث،

$$S_{2i}^2 = \frac{1}{M_i - 1}\sum_{j=1}^{M_i}(y_{ij} - \bar{Y}_i)^2$$

هو التباين بين الوحدات الجزئية في الوحدة.

ويحوي متوسط مربعات خطأ ثلاث مركبات: واحدة تنبثق عن التغير ضمن الوحدات، وواحدة عن التغير بين المتوسطات الحقيقية للوحدات، وواحدة عن الانحياز.

ولم تُحدّد قيم m_i. والاختيار الأكثر شيوعًا هو إما أن نأخذ جميع القيم m_i متساوية، أو نأخذ m_i متناسبة مع M_i أي أن نسبة المعاينة الجزئية من أي وحدة نختارها هي نسبة مثبتة. ولا يؤثر اختيار m_i إلا في المركبة الأولى من المركبات الثلاث للتباين وهي المركبة المنبثقة عن التباين ضمن الوحدات.

II ـ وحدات اختيرت باحتمالات متساوية

$$\text{تقدير} = \bar{y}_{II} = \frac{NM\bar{y}}{M_0}$$

وهذا التقدير غير منحاز وبما أن $\bar{y}_i$ هو تقدير غير منحاز لِـ $\bar{Y}_i$ فإن الجداء $M_i\bar{y}_i$ هو تقدير غير منحاز لمجموع الوحدة Y_i، وبالتالي فإن $NM_i\bar{y}_i$ هو تقدير غير منحاز لمجموع المجتمع Y. وبالقسمة على M_0 العدد الكلي للوحدات الجزئية في المجتمع، نحصل على تقدير غير منحاز لِـ $\bar{\bar{Y}}$.

ولإيجاد $V(\bar{y}_{II})$، وهو بالطبع يساوي متوسط مربعات الخطأ، نجد،

$$\bar{y}_{II} - \bar{\bar{Y}} = \frac{NM_i\bar{y}_i}{M_0} - \bar{\bar{Y}}$$

$$= \frac{NM_i}{M_0}(\bar{y}_i - \bar{Y}_i) + \left(\frac{NM_i}{M_0}\bar{Y}_i - \bar{\bar{Y}}\right)$$

والآن $M_i\bar{Y}_i = Y_i$ أي مجموع الوحدة، و $\bar{\bar{Y}} = N\bar{Y}/M_0$ حيث $\bar{Y}$ متوسط المجتمع لكل وحدة. وهذا يعطي،

$$\bar{y}_{II} - \bar{\bar{Y}} = \frac{NM_i}{M_0}(\bar{y}_i - \bar{Y}_i) + \frac{N}{M_0}(Y_i - \bar{Y})$$

ومنه،

$$V(\bar{y}_{II}) = \frac{N}{{M_0}^2}\sum_{i=1}^{N} M_i(M_i - m_i)\frac{{S_{2i}}^2}{m_i} + \frac{N}{{M_0}^2}\sum_{i=1}^{N}(Y_i - \bar{Y})^2 \qquad (11.2)$$

ومركبة «ما بين الوحدات» لهذا التباين (الحد الثاني من الطرف الأيمن) تمثل التباين بين مجاميع الوحدات Y_i . وتتأثر هذه المركبة بكل من التغيرات في M_i من وحدة إلى وحدة، وبالتغيرات في المتوسطات $\bar{Y}_i$ لكل عنصر. وإذا تغيرت الوحدات بشدّة من حيث حجمها، فإن هذه المركبة تكون كبيرة حتى ولو كانت المتوسطات لكل عنصر $\bar{Y}_i$ ثابتة تقريبًا من وحدة إلى وحدة. وغالبًا ما تكون هذه المركبة كبيرة بحيث يكون لـ $\bar{y}_{II}$ متوسط مربعات خطأ أعلى بكثير من التقدير المنحاز $\bar{y}_I$. وهكذا فإنه لا الطريقة I ولا الطريقة II مُرضية تمامًا.

III ـ وحدات مختارة باحتمالات متناسبة مع الحجم

متوسط العيّنة = $\bar{y}_{III} = \bar{y}_i$ = تقدير

وتعود هذه الطريقة إلى Hansen و Hurwitz (1943) . وهي تعطي متوسط عيّنة غير منحاز وغير خاضع للتضخم الذي يتصف به تباين الطريقة II .

وعند تكرار المعاينة، تظهر الوحدة بتكرار نسبي M_i/M_0 ومنه،

$$E(\bar{y}_{III}) = \sum_{i=1}^{N}\frac{M_i}{M_0}\bar{Y}_i = \bar{\bar{Y}}$$

وبالإضافة إلى ذلك،

$$\bar{y}_{III} - \bar{\bar{Y}} = (\bar{y}_{III} - \bar{Y}_i) + (\bar{Y}_i - \bar{\bar{Y}})$$

لنأخذ أولًا المتوسط فوق العيّنات التي اختيرت فيها الوحدة i

$$E(\bar{y}_{III} - \bar{\bar{Y}})^2 = \left(\frac{M_i - m_i}{M_i}\right)\frac{{S_{2i}}^2}{m_i} + (\bar{Y}_i - \bar{\bar{Y}})^2$$

والآن لنأخذ المتوسط فوق جميع الاختيارات الممكنة للوحدة. وبما أن الوحدة i مختارة بتكرار نسبي M_i/M_0 فإن،

$$V(\bar{y}_{III}) = \frac{1}{M_0}\left[\sum_{i=1}^{N} (M_i - m_i)\frac{S_{2i}^2}{m_i} + \sum_{i=1}^{N} M_i(\bar{Y}_i - \bar{Y})^2\right] \qquad (11.3)$$

وكما في الطريقة I نلاحظ أن مركبة «ما بين الوحدات» تنشأ عن الفروق بين $\bar{Y}_i$ ، «المتوسطات لكل وحدة جزئية» في الوحدات المتتالية. وإذا كانت هذه «المتوسطات لكل وحدة جزئية» متساوية تقريبًا، فتكون هذه المركبة صغيرة.

مثال

لنطبق هذه النتائج على مجتمع صغير، ننشئه اصطناعيًا. المعلومات الإحصائية مقدّمة في الجدول (١١-١)، وتوجد ثلاث وحدات، بـ 2 ، 4 و 6 عناصر، على الترتيب. ويمكن للقارىء التحقق من الأرقام المعطاة لِـ $\bar{Y}_i$ ، Y_i, S_{2i}^2 . ومتوسط المجتمع $\bar{\bar{Y}}$ هو 33/12 أو 2.75 والمتوسط غير المرجّح لِـ $\bar{Y}_i$ هو $\bar{Y}_a = 2.167$ ، بحيث يكون الانحياز في الطريقة I مساوٍ لِـ 0.583 . ومربعه أي المساهمة في متوسط مربعات الخطأ هو 0.340 .

جدول (١١-١) مجتمع اصطناعي بوحدات ذات حجوم غير متساوية

الوحدة	y_{ij}	M_i	Y_i	S_{2i}^2	$\bar{Y}_i$	$\bar{Y}_i - \bar{\bar{Y}}$
1	0, 1	2	1	0.500	0.5	−2.25
2	1, 2, 2, 3	4	8	0.667	2.0	−0.75
3	3, 3, 4, 4, 5, 5	6	24	0.800	4.0	+1.25
	المجاميع	12	33			

وعلينا أن نختار وحدة واحدة ثم نأخذ منها عيّنة من وحدتين جزئيتين. وسنأخذ بعين الاعتبار أربع طرق، اثنتان منهما عبارة عن شكلين مختلفين للطريقة I .

طريقة I_a

الاختيار: وحدة باحتمالات متساوية، $m_i = 2$.

التقدير: $\bar{y}_i$ (منحاز)،

طريقة I_b

الاختيار: وحدة باحتمالات متساوية، $m_i = \frac{1}{2}M_i$

التقدير: $\bar{y}_i$ (منحاز)

طريقة II

الاختيار: وحدة باحتمالات متساوية، $m_i = 2$

التقدير: $NM_i\bar{y}_i/M_0$ (غير منحاز)

طريقة III

الاختيار: وحدة باحتمال M_i/M_0، $m_i = 2$

التقدير: $\bar{y}_i$ (غير منحاز)،

والطريقة *Ib* (معاينة جزئية تناسبية) لا تضمن حجم عيّنة مساوٍ لـ 2 (يمكن أن يكون 1 ، 2 أو 3) ولكن متوسط حجم العيّنة هو 2 .

وبتطبيق علاقات خطأ المعاينة (11.1) ، (11.2) و (11.3) نحصل على النتائج في الجدول (١١-٢) .

جدول (١١-٢) متوسطات مربعات الخطأ لتقديرات عيّنة لـ $\hat{Y}$

الطريقة	المساهمة في متوسط مربعات الخطأ من: ما ضمن الوحدات	ما بين الوحدات	الانحياز	متوسط مربعات الخطأ الكلي
I*a*	0.144	2.056	0.340	2.540
I*b*	0.183	2.056	0.340	2.579
II	0.256	5.792	0.000	6.048
III	0.189	1.813	0.000	2.002

ومع أن المثال مصطنع فالنتائج نسخة تقليدية عن تلك التي وجدت في مقارنات تمت في مجتمعات عديدة. وتعطي الطريقة III أصغر متوسط مربعات خطأ لأننا نعثر في هذه الطريقة على أصغر مساهمة للتغير بين الوحدات. ومع أن الطريقة II غير منحازة إلا أنها أدنى بكثير والطريقة (حجوم متساوية للعيّنات الجزئية) أفضل بقليل من الطريقة I_b (معاينة جزئية تناسبية) .

وقد تمت بعض المقارنات بين هذه الطرق في مجتمعات واقعية. ومن أجل ست مفردات (العدد الكلي للعمال، العدد الكلي للعمال الزراعيين، العدد الكلي للعمال غير الزراعيين، مقدرة بصورة منفصلة من أجل الذكور والإناث) وجد Hansen و Hurwitz (1943) أن الطريقة III قد أنتجت تخفيضات كبيرة في مساهمة التغير بين الوحدات وذلك بالمقارنة مع الطريقة II غير المنحازة، وتخفيضات كانت في متوسطها 30 بالمائة بالمقارنة مع الطريقة I . (افترضوا أن مساهمة التغير ضمن الوحدات مهملة). وقد ذكر Jebe (1952) عند تقديره لمفردات زراعية تقليدية في ولاية نورث كارولاينا، أن التخفيضات في التباين الكلي كانت من مرتبة 15 بالمائة بالمقارنة مع طرق من النوع I . وكانت المنطقة هي الوحدة الأولية في كل من الدراستين.

(١١-٣) المعاينة مع احتمالات متناسبة مع الحجم المقدّر

كما ذكرنا في الفصل التاسع تكون حجوم الوحدات M_i مأخوذة أحيانًا من بيانات سابقة، وبالتالي فهي معروفة بصورة تقريبية فقط. وفي مسوح إحصائية أخرى قد تتوافر عدة قياسات ممكنة لحجم الوحدة. ليكن z_i الاحتمال أو الحجم النسبي المخصص للوحدة i ، حيث الأعداد z_i أي مجموعة من الأعداد الموجبة مجموعها الواحد. ولا نزال نفترض $n=1$.

طريقة *IV*

التقدير

$$\bar{y}_{\text{IV}} = \frac{M_i \bar{y}_i}{z_i M_0} \tag{11.4}$$

هو تقدير غير منحاز لـ $\bar{Y}$ ذلك لأنه عند تكرار المعاينة، تظهر الوحدة i بتكرار نسبي z_i ، بحيث يكون

$$E(\bar{y}_{\text{IV}}) = \sum_{i=1}^{N} z_i \left(\frac{M_i \bar{y}_i}{z_i M_0} \right) = \sum_{i=1}^{N} \frac{M_i \bar{y}_i}{M_0} = \bar{Y}$$

ونحصل على تباين $\bar{y}_{IV}$ بالطريقة العادية . لنكتب،

[بالاستناد إلى (11.4)]

$$\bar{y}_{IV}-\bar{\bar{Y}}=\frac{M_i\bar{y}_i}{z_iM_0}-\bar{\bar{Y}}$$

$$=\frac{1}{M_0}\left[\frac{M_i}{z_i}(\bar{y}_i-\bar{Y}_i)+\left(\frac{M_i}{z_i}\bar{Y}_i-M_0\bar{\bar{Y}}\right)\right]$$

وفي عبارة التباين يتلقى كل مربع الترجيح z_i ومنه :

$$V(\bar{y}_{IV})=\frac{1}{M_0^2}\left[\sum_{i=1}^{N}\frac{M_i(M_i-m_i)}{z_i}\frac{S_{2i}^2}{m_i}+\sum_{i=1}^{N}z_i\left(\frac{M_i\bar{Y}_i}{z_i}-M_0\bar{\bar{Y}}\right)^2\right] \qquad (11.5)$$

وإذا كان $z_i=M_i/M_0$ فإن العلاقــة (11.5) تختــزل إلى العــلاقــة (11.3) المــوافقــة لـ $V(\bar{y}_{III})$. وإذا كان $z_i=1/N$ (الاحتمالات الابتــدائية متســاوية) فإن العــلاقــة (11.5) تُختــزل إلى العــلاقــة (11.2) المــوافقــة لتبــاين تقــدير غير منحاز عندما تكون الاحتمالات متساوية .

وما لم يكن $z_i=M_i/M_0$ فإن مركبة «ما بين الوحدات» في (11.5) تتأثر إلى حد ما بالتغير في الحجوم M_i . كما تتأثر بالتغيرات في المتوسطات $\bar{Y}_i$ لكل عنصر .

جدول (١١-٣) حساب $V(\bar{y}_{IV})$

الوحدة	M_i	M_i/M_0	z_i	m_i	$\frac{M_i(M_i-m_i)}{z_im_i}$	S_{2i}^2	Y_i	$\frac{Y_i}{z_i}$	$\frac{Y_i}{z_i}-Y$
1	2	0.17	0.2	2	0	0.500	1	5	−28
2	4	0.33	0.4	2	10	0.667	8	20	−13
3	6	0.50	0.4	2	30	0.800	24	60	+27

مثــال

يبــيّن الجــدول (١١-٣) الحســابــات الأســاسية لإيجــاد $V(\bar{y}_{IV})$ في المجتمــع الاصطناعي المبيّن في الجدول (١١-١) . وقد أخذت z_i على أنها 0.2 ، 0.4 و $m_i=2$.
ومن العلاقة (11.5) نجد التباين كما يلي :

$$\text{مساهمة «ماضمن الوحدات»} = \sum \frac{M_i(M_i - m_i)S_{2i}^2}{z_i m_i} / M_0^2 = 0.213$$

$$\text{مساهمة «مابين الوحدات»} = \sum z_i\left(\frac{Y_i}{z_i} - Y\right)^2 / M_0^2 = 3.583$$

وتكشف المقارنة مع الجدول (١١-٢) أن للطريقة IV تبايناً أقل من تباين الطريقة غير المنحازة II التي نختار فيها الوحدات الأولية باحتمالات متساوية، ولكن الطريقة IV هي بلا ريب مختلفة عن الطريقة I أو الطريقة III . وتدفع الطريقة IV في هذا المثال ثمناً مرتفعاً جدًّا لقاء الحصول على تقدير غير منحاز. وبالتالي فمن الطبيعي أن ندرس ما إذا كان يمكن لمتوسط العيّنة (كما في الطريقة I) أن يكون أفضل من التقدير المتبنّى في الطريقة IV .

V ـ *وحدات مختارة باحتمالات متناسبة مع تقدير الحجم*

$$\text{تقدير} = \bar{y}_V = \bar{y}_i = \text{متوسط العيّنة}$$

والتقدير منحاز باعتبار أنه، على سبيل المثال،

$$E(\bar{y}_i) = \sum z_i \bar{Y}_i = \bar{Y}_z$$

وإذا كانت z_i تقديرات جيدة، فإن $\bar{Y}_z$ تكون قريبة من المتوسط الصحيح $\bar{Y} = \sum M_i \bar{Y}_i / M_0$ والانحياز صغير.

وإذا كتبنا:

$$\bar{y}_V - \bar{Y} = (\bar{y}_i - \bar{Y}_i) + (\bar{Y}_i - \bar{Y}_z) + (\bar{Y}_z - \bar{Y})$$

فالمركبات الثلاث لمتوسط مربعات الخطأ كما يلي:

$$\text{MSE}(\bar{y}_V) = \sum_{i=1}^{N} \frac{z_i(M_i - m_i)}{M_i}\frac{S_{2i}^2}{m_i} + \sum_{i=1}^{N} z_i(\bar{Y}_i - \bar{Y}_z)^2 + (\bar{Y}_z - \bar{Y})^2 \qquad (11.6)$$

حيث ترمز MSE لمتوسط مربعات الخطأ.

مثـــال

إذا اختيرت قيم z_i و m_i كما في الجدول (١١-٣)، فيمكن للقارىء التحقق من أن مركبات تباين $\bar{y}_V$ هي كما تظهر في الجدول (١١-٤).

جدول (١١-٤) المساهمات في متوسط مربعات الخطأ عند تطبيق الطريقة ٧

ضمن الوحدات	بين الوحدات	انحياز	الكلي MSE
0.173	1.800	0.062	2.035

وتتفوق هذه الطريقة على جميع الطرق باستثناء الطريقة III (احتمالات متناسبة مع الحجم) وهي تقريبًا في جودة الطريقة III نفسها.

(١١-٤) تلخيص للطرق في حالة n=1

ويلخص الجدول (١١-٥) الطرق الخمس لتقدير المتوسط لكل عنصر $\bar{Y}$ ومتوسطات خطئها التربيعي الواردة في المثال العددي أعلاه.

ولتقدير مجموع المجتمع $Y = M_0\bar{Y}$. تُضرب التقديرات السابقة بـ M_0 ومتوسطات خطئها التربيعي في الجدول (١١-٥) بـ$M_0^2 = 144$ وإنجازاتها النسبية تبقى نفسها.

جدول (١١-٥) طرق المعاينة على مرحلتين $(n=1)$

الطريقة	احتمالات اختيار الوحدات	تقدير $\bar{Y}$	واقع الانحياز	MSE في المثال
I	متساوية	$\bar{y}_i$	منحاز	Ia: 2.541 Ib: 2.579
II	متساوية	$\frac{NM_i\bar{y}_i}{M}$	غير منحاز	6.048
III	الحجم $\propto \frac{M_i}{M_0}$	$\bar{y}_i$	غير منحاز	2.002
IV	تقدير الحجم $\propto z_i$	$\frac{M_i\bar{y}_i}{z_iM_0}$	غير منحاز	3.796
V	تقدير الحجم $\propto z_i$	$\bar{y}_i$	منحاز	2.035

(١١-٥) طرق المعاينة في حالة n>1

في حالة $n>1$ نجد أن طرق المعاينة الرئيسة تعميمات طبيعية للطرق السابقة I إلى IV وللطرق التي نوقشت في الفصل ١٩ في حالة معاينة وحيدة المرحلة مع

وحدات عنقودية ذات حجوم غير متساوية لكل وحدة جزئية، والمتوسطات أو النسب التي لها بنية التقديرات النسبة.

وفي العـديد من التطبيقات لا نعلم الكمية M_0 ، أي العدد الكلي للوحدات الجزئية في المجتمع، ولكننا نعلم M_i فقط في الوحدات الأولية التي سحبت في العيّنة باعتبار أنه يمكن تعداد هذه الكميات M_i من خلال قوائم جاهزة. ومما يجدر ذكره إذن أن الـطريقتين II و IV وتعميمها إلى حالة $n > 1$ لا تتطلبان معرفة M_0 من أجل تقدير مجموع المجتمع. ولا تحتاج الطريقة I وتعميمها إلى معرفة M_0 من أجل تقدير المتوسط لكـل وحـدة جزئية. أمـا الـطريقـة III ، احتـمالات متناسبة مع الحجم، فتتطلب معرفة M_0.

وتدعى التقديرات ذاتية الترجيح، عندما تكون ببساطة من مضاعفات المجموع فوق جميع الوحدات الجزئية في العيّنة. وبالنظر إلى السهولة التطبيقية للتقديرات ذاتية الترجيح في مسوح إحصائية تتم على نطاق واسع، فسنذكر الشرط الذي يصبح معه كل تقدير تقديرًا ذاتي الترجيح. والشرط في حالة تقديرات غير منحازة هو، كما سنرى، أن تتـوافـر فرص سحب أو اختيار متسـاوية لكـل وحدة من وحدات المرحلة الثانية في المجتمع، أو بصورة أعم لكل وحدة من وحدات المرحلة النهائية في المعاينة.

(١١-٦) نتيجتان مفيدتان

هناك نتيجتان مفيدتان عند إيجاد تباينات وتقديرات العيّنة. وهما تبيّنان كيفية تعميم نتـائـج المعاينة وحيدة المرحلة إلى معاينة بمرحلتين أو بصورة أعم إلى معاينة متعددة المراحل. وأول من برهنها كان Durbin (1953) لمقدرات هي مجاميع مقدرات تتعلق بالـوحدات الأولية في العيّنة، وقد عمّمها Des Raj (1966) إلى دوال خطية في مقدرات الوحدات الأولية، كما عممها Rao (1975 b) إلى حالات أكثر تعقيدًا. وسنتبع الطريقة التي استخدمها Des Raj .

ويجـري اختيار الـوحـدات الأولية بدون إعـادة، باحتـمالات متساوية أو غير متسـاوية. وفي الـوحـدة الأولية i ، ليكن $\hat{Y}_i$ مقـدّرًا غير منحـاز لمجمـوع الـوحـدة

Y_i ، مع تباين مرحلة ثانية σ_{2i}^2 وتكون المعاينة الجزئية مستقلة في الوحدات الأولية المختلفة. لنأخذ مقدرًا غير منحاز لمجموع المجتمع Y من الشكل،

$$\hat{Y} = \sum_{i=1}^{n} w_{is} \hat{Y}_i \tag{11.7}$$

حيث الترجيحة w_{is} معروفة في كل عيّنة s . وقد تعتمد على وحدات أولية أخرى موجودة في العيّنة كما تعتمد أيضًا على الوحدة i .

وسنتبنّى الحيلة المستخدمة في الفقرة (٢-٩) فنعتبر w_{is}' متغيراً عشوائيًا يساوي w_{is} إذا ظهرت الوحدة i في العيّنة ويساوي الصفر فيما عدا ذلك. وهذا يعطي،

$$\hat{Y} = \sum_{i=1}^{N} w_{is}' \hat{Y}_i \tag{11.8}$$

والآن

$$E(\hat{Y}) = E_1 E_2(\hat{Y}) = E_1\left(\sum_{i=1}^{N} w_{is}' Y_i\right) = Y \tag{11.9}$$

إذا وفقط إذا كان لكل $E_1(w_{is}') = 1$ لكل i .

نظرية (١١-١)

$$V(\hat{Y}) = V\left(\sum_{i=1}^{n} w_{is} \hat{Y}_i\right) = V\left(\sum_{i=1}^{n} w_{is} Y_i\right) + \sum_{i=1}^{N} E_1(w_{is}'^2)\sigma_{2i}^2 \tag{11.10}$$

برهان

من العلاقة (10.2) نجد،

$$\begin{aligned} V(\hat{Y}) &= V_1[E_2(\hat{Y})] + E_1[V_2(\hat{Y})] \\ &= V\left(\sum_{i=1}^{n} w_{is} Y_i\right) + E_1\left[\sum_{i=1}^{N} w_{is}'^2 V_2(\hat{Y}_i)\right] \end{aligned} \tag{11.11}$$

وبما أن تغاير المرحلة الثانية بين $\hat{Y}_i$ و $\hat{Y}_j$ $(i \neq j)$ هو الصفر، باعتبار أن المعاينة الجزئية مستقلة ومنه،

$$V(\hat{Y}) = V\left(\sum_{i=1}^{n} w_{is} Y_i\right) + \sum_{i=1}^{N} [E_1(w_{is}'^2)\sigma_{2i}^2] \tag{11.12}$$

وهو المطلوب.

والحد الأول في (11.12) هو تباين المقدِّر الموافق عندما نحصي بالكامل كل وحدة أولية في العيّنة، ونحصل عليه من نتائج المعاينة على مرحلة واحدة.

مثـال

من أجل المقدّر المماثل لمقدّر Horvitz-Thompson في حالة معاينة على مرحلتين، نجد، $\hat{Y}_{HT} = \sum^{n} \hat{Y}_i/\pi_i$ والترجيحة هي $w_{is}' = 1/\pi_i$ إذا كانت الوحدة i في العيّنة وصفر فيما عدا ذلـك. وبالتالي فإن $E_1(w_{is}'^2) = \pi_i/\pi_i^2 = 1/\pi_i$ حيث π_i احتمال سحب الوحدة i. وفضلاً عن ذلك، إذا سحبنا، بمعاينة عشوائية بسيطة m_i وحدة جزئية من الـ M_i، وذلك حيثما تُسحب الوحدة i، فعندئذ،

$$\sigma_{2i}^2 = V_2(\hat{Y}_i) = \frac{M_i(M_i - m_i)}{m_i} S_{2i}^2 \tag{11.13}$$

وبالتالي، واستنادًا إلى النظرية (١١-١) مستخدمين العلاقة (9A.42) الخاصة بالتباين المقابل $V(\hat{Y}_{HT})$ في حالة معاينة وحيدة المرحلة، نجد،

$$V(\hat{Y}_{HT}) = \sum_{i=1}^{N}\sum_{j>i}^{N}(\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 + \sum_{i=1}^{N}\frac{M_i(M_i - m_i)S_{2i}^2}{m_i\pi_i} \tag{11.14}$$

نظرية (١١-٢)

لنفرض أن لدينا تقديرًا غير منحاز $\hat{\sigma}_{2i}^2$ لِـ σ_{2i}^2 وهو تباين المرحلة الثانية لـ $\hat{Y}_i$ وتقدير عيّنة غير منحاز لِـ $V(\sum^{n} w_{is} Y_i) = V(\sum^{N} w_{is}' Y_i)$ من معاينة المرحلة الأولى. فسيكون هذا الأخير صيغة تربيعية من النوع،

$$v\left(\sum^{n} w_{is} Y_i\right) = \sum_{i}^{n} a_{is} Y_i^2 + 2\sum_{i}^{n}\sum_{j>i}^{n} b_{ijs} Y_i Y_j \tag{11.15}$$

وعندئذ يكون،

$$v\left(\sum^{n} w_{is}\hat{Y}_i\right) = \left(\sum_{i}^{n} a_{is}\hat{Y}_i^2 + 2\sum_{i}^{n}\sum_{j>i}^{n} b_{ijs}\hat{Y}_i\hat{Y}_j\right) + \sum^{n} w_{is}\hat{\sigma}_{2i}^2 \tag{11.16}$$

تقدير عيّنة غير منحاز لِـ $V(\sum^n w_{is}\hat{Y}_i)$ وهكذا تكون قاعدة وضع مقدّر العيّنة لِـ $V(\sum w_{is}Y_i)$ كما يلي: في مقدّر عيّنة غير منحاز لِـ $V(\sum^n w_{is}Y_i)$ من معاينة المرحلة الأولى، ضع $\hat{Y}_i$ بدلاً من Y_i وذلك حيثما ظهرت Y_i، ثم أضف إلى هذا الحدّ $\sum^n (w_{is}\hat{\sigma}_{2i}{}^2)$ حيث $\hat{Y} = \sum^n w_{is}\hat{Y}_i$ و $\hat{\sigma}_{2i}{}^2$ مقدّر غير منحاز لِـ $V_2(\hat{Y}_i)$.

برهان

$$V\left(\sum_i^N w_{is}'Y_i\right) = \sum_i^N Y_i^2 V(w_{is}') + 2\sum_i^N \sum_{j>i}^N Y_iY_j \operatorname{cov}(w_{is}'w_{js}') \qquad (11.17)$$

أدخل المتغير العشوائي، a_{is}' حيث $a_{is}' = a_{is}$ إذا كانت الوحدة ضمن العيّنة و $a_{is}' = 0$ فيما عدا ذلك. وبصورة مشابهة، ليكن $b'_{ijs} = b_{ijs}$ إذا كانت الوحدتان i و j ضمن العيّنة و $b'_{ijs} = 0$ فيما عدا ذلك. ومن (11.5) في حالة معاينة وحيدة المرحلة نجد،

$$v\left(\sum^N w_{is}'Y_i\right) = \sum_i^N a_{is}'Y_i^2 + 2\sum_i^N \sum_{j>i}^N b'_{ijs}Y_iY_j \qquad (11.18)$$

وإذا أردنا لهذا أن يكون غير منحاز فإن مقارنة (11.18) و (11.17) **تبيّن أن** $E_1(a_{is}')$ يجب أن يكون مساويًا لِـ $V(w_{is}')$ والآن نجد من أجل مقدر التباين (11.16)،

$$E_1E_2\left(\sum_i^N a_{is}'\hat{Y}_i^2 + 2\sum_i^N \sum_{j>i}^N b'_{ijs}\hat{Y}_i\hat{Y}_j\right) + E_1E_2\left(\sum^N w_{is}'\hat{\sigma}_{2i}'^2\right)$$

$$= E_1\left(\sum_i^N a_{is}'Y_i^2 + 2\sum_i^N \sum_{j>i}^N b'_{ijs}Y_iY_j\right) + \sum_i^N [V(w_{is}') + E^2(w_{is}')]\sigma_{2i}{}^2 \qquad (11.19)$$

وقد استخدمنا في (11.19) النتيجتين، $E(a_{is}') = V(w_{is}')$ وأن $E_1(w_{is}') = 1 = E_1{}^2(w_{is}')$ مهما تكن قيمة i . ومنه، بالاستناد إلى (11.12) يكون

$$E\left[v\left(\sum^n w_{is}\hat{Y}\right)\right] = V\left(\sum^n w_{is}Y_i\right) + \sum_i^N E_1(w_{is}'^2)\sigma_{2i}{}^2 = V\left(\sum_i^n w_{is}\hat{Y}_i\right) \qquad (11.20)$$

وهو المطلوب.

لنعتبر الآن بعض المقدّرات المحددة لمجموع المجتمع Y . ففي المسوح الإحصائية المتسعة أصبح اختيار الوحدات الأولية باحتمالات غير متساوية الطريقة

الأكثر شيوعًا. وتعالج الفقرتان (١١-٩) (المعاينة مع الإعادة) و(١١-١٠) (المعاينة بدون إعادة) مثل هذه الطرق، وتقدم الفقرة (١١-١٣) المقدّر النسبة في معاينة (أ م مح)*. والفقرات الأخرى (١١-٧) و(١١-٨) تعطي الطرق الموافقة عندما يتم اختيار الوحدات الأولية باحتمالات متساوية.

(١١-٧) وحدات اختيرت باحتمالات متساوية ـ مقدّر غير منحاز

ما لم نذكر غير ذلك، نفترض أن اختيار وحدات العيّنة الجزئية الـ m_i من الوحدة i قد جرى وفق معاينة عشوائية بسيطة. والمقدّر غير المنحاز لمجموع المجتمع هو،

$$\hat{Y}_u = \frac{N}{n}\sum^{n} M_i\bar{y}_i = \frac{N}{n}\sum^{n} \hat{Y}_i \tag{11.21}$$

ولتطبيق النظرية (١١-١)، نلاحظ أن،

$$w_{is} = \frac{N}{n}; \qquad E(w_{is}{'}) = \frac{n}{N}\frac{N}{n} = 1; \qquad E(w_{is}{'^2}) = \frac{n}{N}\frac{N^2}{n^2} = \frac{N}{n}$$

وبالتالي نجد من النظرية (١١-١) والنتائج التي تسبقها أن،

$$V(\hat{Y}_u) = \frac{N^2}{n}(1-f_1)\frac{\sum(Y_i-\bar{Y})^2}{(N-1)} + \frac{N}{n}\sum^{N}\frac{M_i^2(1-f_{2i})S_{2i}{^2}}{m_i} \tag{11.22}$$

حيث $f_{2i} = m_i/M_i$ ويصبح المقدّر ذاتي الترجيح إذا كان f_{2i} ثابتًا (يساوي f_2 مثلاً). ولدينا عندئذ،

$$\hat{Y}_u = \frac{N}{nf_2}\sum_{i}^{m}\sum_{j}^{m_i} y_{ij} \tag{11.23}$$

والكمية nf_2/N هي، بالطبع، احتمال سحب أي وحدة من وحدات المرحلة الثانية.

ومن أجل تقدير عيّنة غير منحاز للتباين، تعطي النظرية (١١-٢)، بالاستفادة من (11.6)،

---

* أم مح ترمز لاحتمالات متناسبة مع مقدار يمكن اعتباره ممثلاً للحجم. ويُقابلها بالإنكليزية PPZ.

المترجم

$$v(\hat{Y}_u)=\frac{N^2(1-f_1)}{n}\frac{\sum_i^n(\hat{Y}_i-\hat{\bar{Y}}_u)^2}{n-1}+\frac{N}{n}\sum_i^n\frac{M_i^2(1-f_{2i})s_{2i}^2}{m_i} \qquad (11.24)$$

(١١ـ ٨) وحدات اختيرت باحتمالات متساوية: تقدير نسبة إلى الحجم

هذا المقدّر لمجموع المجتمع Y هو،

$$\hat{Y}_R=M_0\frac{\sum^n M_i\bar{y}_i}{\sum^n M_i}=M_0\frac{\sum^n \hat{Y}_i}{\sum^n M_i} \qquad (11.25)$$

وهو المقدّر النسبة التقليدي، باعتبار أن كلًّا من البسط والمقام يتغير من عيّنة إلى عيّنة. ويستخدم بصورة رئيسة لتقدير المتوسطات على أساس الوحدة الجزئية، وهي متوسطات لا نحتاج فيها لمعرفة M_0، ويمثل تعميمًا للطريقة I مع $n=1$.

ولإيجاد متوسط خطئه التربيعي بصورة تقريبية، نكتب،

$$\hat{Y}_R-Y=M_0\frac{\sum^n M_i\bar{y}_i}{\sum^n M_i}-Y\doteq\frac{N}{n}\sum^n M_i(\bar{y}_i-\bar{\bar{Y}}) \qquad (11.26)$$

حيث $\bar{\bar{Y}}=Y/M_0$.

وبما أن $\hat{Y}_u=(N/n)\sum^n M_i\bar{y}_i$ فيمكن الحصول على $\text{MSE}(\hat{Y}_R)$ التقريبي من العلاقة (11.22) الخاصة بـ $V(\hat{Y}_u)$ وذلك بوضع $M_i(\bar{y}_i-\bar{\bar{Y}})$ بدلاً من $M_i\bar{y}_i$، أو بصورة أعم، وضع $(y_{ij}-\bar{\bar{Y}})$ بدلاً من y_{ij} والآن (11.22) تكون،

$$V(\hat{Y}_u)=\frac{N^2}{n}(1-f_1)\frac{\sum^N(Y_i-\bar{Y})^2}{N-1}+\frac{N}{n}\sum^n\frac{M_i^2(1-f_{2i})S_{2i}^2}{m_i}$$

وعند التعويض، تصبح $Y_i=M_i\bar{Y}_i$ على الشكل $M_i(\bar{Y}_i-\bar{\bar{Y}})$ بينما نضع صفرًا بدلاً من $\bar{Y}=\sum^N y_i/N$. وبالإضافة إلى ذلك، يبقى $S_{2i}^2=\sum_j(y_{ij}-\bar{Y}_i)^2/(M_i-1)$ بدون تغيير عند وضع $(y_{ij}-\bar{\bar{Y}})$ بدلاً من y_{ij} وهذا يعطي النتيجة،

$$\text{MSE}(\hat{Y}_R)\doteq\frac{N^2}{n}(1-f_1)\frac{\sum^N M_i^2(\bar{Y}_i-\bar{\bar{Y}})^2}{N-1}+\frac{N}{n}\sum^N\frac{M_i^2(1-f_{2i})S_{2i}^2}{m_i} \qquad (11.27)$$

وكما في حالة $\hat{Y}_\theta$ يصبح هذا المقدّر ذاتي الترجيح إذا كان،

$$f_{2i} = \frac{m_i}{M_i} = \text{ثابت} = f_2 = \frac{\bar{m}}{\bar{M}} = \frac{N\bar{m}}{M_0}$$

وفي هذه الحالة يمكن التعبير عن مساهمة ما ضمن الوحدات ببساطة أكثر، وذلك بوضع $m_i = N\bar{m}M_i/M_0$ مما يعطي،

$$\text{MSE}(\hat{Y}_R) \doteq \frac{N^2}{n}(1-f_1)\frac{\sum^N M_i^2(\bar{Y}_i - \bar{\bar{Y}})^2}{(N-1)} + \frac{M_0^2(1-f_2)}{n\bar{m}}\sum^N\left(\frac{M_i}{M_0}\right)S_{2i}^2 \quad (11.28)$$

ويمكن ملاحظة الشبه بالعلاقة الموافقة عندما تكون الوحدات الأولية ذات حجوم متساوية. ومن (10.14) فقرة (١٠-٣) وباعتبار $\hat{Y} = M_0\bar{\bar{y}}$ نجد بعد الضرب بـ M_0^2،

$$V(\hat{Y}) = \frac{M_0^2(1-f_1)}{n}\frac{\sum^N(\bar{Y}_i - \bar{\bar{Y}})^2}{(N-1)} + \frac{M_0^2(1-f_2)}{nm}\sum^N\left(\frac{1}{N}\right)S_{2i}^2 \quad (11.29)$$

والفرق هو أنه في (11.28) تكون مساهمات الوحدات الأولية في متوسط الخطأ التربيعي مساهمات مرجّحة، وتتلقى الوحدات الأكبر ترجيحة أكبر.

وبالاستناد إلى النظرية (١١-٢) نعطي هنا تقدير عيّنة تقريبي لِـ $\text{MSE}(\hat{Y}_R)$ المذكور في (11.27) وهو،

$$v(\hat{Y}_R) = \frac{N^2}{n}(1-f_1)\frac{\sum^n M_i^2(\bar{y}_i - \hat{\bar{\bar{Y}}}_R)^2}{n-1} + \frac{N}{n}\sum^n\frac{M_i^2(1-f_{2i})s_{2i}^2}{m_i} \quad (11.30)$$

حيث $\hat{\bar{\bar{Y}}}_R = \sum^n M_i\bar{y}_i/\sum^n M_i = \hat{Y}_R/M_0$.

وعند استخدام هذا المقدّر لتقدير متوسط المجتمع لكل وحدة جزئية نجد $\hat{\bar{\bar{Y}}}_R = \sum^n \hat{Y}_i/\sum^n M_i$ و $v(\hat{\bar{\bar{Y}}}_R) = v(\hat{Y}_R)/M_0^2$ وعندما يكون M_0 غير معروف، نعوّض M_0 بِـ $\hat{M}_0 = N(\sum^n M_i/n)$ عند حساب $v(\hat{\bar{\bar{Y}}}_R)$

وعند اختيار الوحدات الأولية باحتمالات متساوية فإن تقديرًا بديلاً لمتوسط المجتمع لكل وحدة جزئية (تعميم آخر للطريقة I) هو،

$$\frac{1}{n}(\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_n)$$

وهذا التقدير ذاتي الترجيح إذا كان m_i ثابتًا، كما في المثال التالي. وعندما لا يوجد ارتباط بين m_i و $\bar{Y}_i$ يمكن أن يكون هذا التقدير مُرضيًا، إلا إنه عرضة للانحياز الذي لا ينعدم حتى في حالة n كبير.

مثـــال

اختـيرت 20 صفحـة عشـوائيًا من مجلد «رجال العلم الأمريكيين». ومن كل صفحة اختيرت عشوائيًا سيرتان، وسجلنا عمر العالم في كل منهما. ويختلف العدد الكلي للسّير في الصفحة الواحدة من حوالي 14 إلى 21 . قدّر متوسط العمر وانحرافه المعياري من البيان المعطى في الجدول (١١-٦) مستخدمًا التقدير النسبة.

جدول (١١-٦) أعمار أربعين عالمًا من رجال العلم الأمريكيين $(m=2,n=20)$

رقم الوحدة	M_i	الأعمار y_{i1}	y_{i2}	المجاميع y_i	$M_i\bar{y}_i$
1	15	47	30	77	577.5
2	19	38	51	89	845.5
3	19	43	45	88	836.0
4	16	55	41	96	768.0
5	16	59	45	104	832.0
6	19	39	38	77	731.5
7	18	43	43	86	774.0
8	18	49	51	100	900.0
9	18	45	35	80	720.0
10	18	46	59	105	945.0
11	20	71	64	135	1,350.0
12	18	35	46	81	729.0
13	19	61	54	115	1,092.5
14	19	45	87	132	1,254.0
15	18	31	38	69	621.0
16	16	64	39	103	824.0
17	16	63	47	110	880.0
18	19	36	33	69	655.5
19	19	61	39	100	950.0
20	19	54	34	88	836.0
المجموع	359			1,904	17,121.5

ومن العمود في أقصى اليمين نجد،

$$\hat{Y}_R = \frac{\sum M_i \bar{y}_i}{\sum M_i} = \frac{17,121.5}{359} = 47.7 \quad \text{سنة}$$

وبما أن n/N مهملة، نجد من (11.30) بعد القسمة على

$$v(\hat{Y}_R) \doteq \frac{\sum^n M_i^2(\bar{y}_i - \hat{Y}_R)^2}{n\bar{M}^2(n-1)}$$

وقد حسبنا البسط على الشكل:

$$\sum(M_i\bar{y}_i)^2 - 2\hat{Y}_R \sum (M_i\bar{y}_i)M_i + \hat{Y}_R{}^2 \sum M_i^2$$
$$= 15,375,020 - (95.3844)(309,747.5) + (2274.55)(6481) = 571,300$$

وبما أن $\bar{M}_n = 359/20$ كما قُدّرت من العيّنة، فهذا يعطي،

$$v(\hat{Y}_R) = \frac{(20)(571,300)}{(19)(359)^2} = 4.67$$
$$s(\hat{Y}_R) = 2.16 \quad \text{سنـة}$$

(١١-٩) وحدات اختيرت باحتمالات غير متساوية مع الإعادة ـ مقدّر غير منحاز

اختيرت الوحدات الأولية باحتمالات متناسبة مع z_i مع الإعادة. وتتبع النتائج في حالة $z_i = M_i/M_0$ (احتمال متناسب مع الحجم) كحالة خاصة.

ويُفترض أن العيّنة الجزئية من m_i وحدة جزئية من الوحدة i قد سُحبت عشوائيًا بدون إعادة. وإذا اختيرت الوحدة i أكثر من مرة، نفترض عند كل اختيار أن مجمل العيّنة الجزئية قد استُبدلت، وأن سحبًا جديدًا مستقلاً لِـ m_i من الوحدات قد جرى بدون إعادة من الوحدة بكاملها.

ويكون،

$$\hat{Y}_{ppz} = \frac{1}{n}\sum^n \frac{M_i\bar{y}_i}{z_i} = \frac{1}{n}\sum^n \frac{\hat{Y}_i}{z_i} \tag{11.31}$$

تقديرًا غير منحاز لمجموع المجتمع.

وفي حالة $n=1$ برهنا في الفقرة (١١-٣) أن هذا المقدّر $M_0\bar{y}_{IV} = \hat{Y}_{IV}$ غير منحاز. ونحصل على تباينه من العلاقة (11.5) لدى ضربها بِـ $M_0{}^2$،

$$V(\hat{Y}_{IV}) = \sum_{i=1}^{N} z_i\left(\frac{Y_i}{z_i} - Y\right)^2 + \sum_{i=1}^{N} \frac{M_i(M_i - m_i)S_{2i}^2}{z_i m_i} \tag{11.32}$$

وبهـذه الـطريقـة في المعاينة نجد أن المقدّر $\hat{Y}_{ppz}$ هو متوسط n من التقديرات المستقلة من الشكـل $\hat{Y}_{IV}$. وبـالتـالي، ومن النـظرية الكلاسيكية في المعاينة يكون $\hat{Y}_{ppz}$ غير منحاز ويكون،

$$V(\hat{Y}_{ppz}) = \frac{1}{n} V(\hat{Y}_{IV}) = \frac{1}{n} \sum_{i=1}^{N} z_i\left(\frac{Y_i}{z_i} - Y\right)^2 + \frac{1}{n} \sum_{i=1}^{N} \frac{M_i^2(1-f_{2i})S_{2i}^2}{m_i z_i} \tag{11.33}$$

وفضلاً عن ذلك، إذا كان لدينا n من التقديرات المستقلة $\hat{Y}_{IV} = Y_i/z_i$، فيكون،

$$v(\hat{Y}_{IV}) = \frac{\sum_{i=1}^{n}\left(\frac{\hat{Y}_i}{z_i} - \hat{Y}_{ppz}\right)^2}{(n-1)} \tag{11.34}$$

بالطبع، مقدّر عيّنة غير منحاز لِـ $V(\hat{Y}_{IV})$

وبالتالي تكون العبارة البسيطة جدًّا،

$$v(\hat{Y}_{ppz}) = \frac{\sum_{i=1}^{n}\left(\frac{\hat{Y}_i}{z_i} - \hat{Y}_{ppz}\right)^2}{n(n-1)} \tag{11.35}$$

مقدّر عيّنة غير منحاز لِـ $V(\hat{Y}_{ppz})$.

وتصـح هذه النتـائـج أيضًـا في معـاينـة متعـددة المراحل، شريطة أن يكون $\hat{Y}_i$ مقدرًا غير منحاز لِـ Y_i ، وأن المعاينة الجزئية مستقلة حيثما تُسحب وحدة أولية.

ولاكتشاف متى تصبح $\hat{Y}_{ppz}$ ذاتية الترجيح في معاينة على مرحلتين، نكتب،

$$\hat{Y}_{ppz} = \frac{1}{n} \sum_{i}^{n} \frac{M_i}{z_i m_i} \sum_{j}^{m_i} y_{ij} \tag{11.36}$$

ويكون الشرط إذن،

$$\frac{n z_i m_i}{M_i} = \text{ثابت} = f_0 \tag{11.37}$$

وهـذه العبـارة هي احتـمال سحب أي وحـدة مرحلة ثانية محددة. ومن (11.37) نجد $m_i/M_i = f_0/nz_i$. وإذا اختير الاحتمال f_0 سلفًا، فيمكن إخبار العامل الميداني مقدّمًا عن كسر المرحلة الثانية m_i/M_i الذي يأخذه من أي وحدة أولية جرى اختيارها. وعلى سبيل المثال، لنفرض أن $f_0=1/50=0.02$ وأن $n=60$ وحدة أولية قد جرى اختيارها. فإذا كان $z_i=0.026$ من أجل وحـدة اخـتيرت فيجـب أن يكـون $m_i/M_i=0.02/(60)(0.0026)$ أو 1 من 7.8 .

ومع عيّنة ذاتية الترجيح، يأخذ مقدّر التباين (11.35) الشكل الأبسط،

$$v(\hat{Y}_{ppz}) = \frac{n}{(n-1)f_0^2}\sum_i^n (y_i - \bar{y})^2 \tag{11.38}$$

حيث $y_i = \sum_j y_{ij}$ مجمـوع العيّنـة من الـوحـدة i وبسـاطـة التبـاينـات المقدّرة (11.35) و (11.38) هي خاصة جذّابة للمعاينة مع الإعادة.

ومع هذه الطريقة توجد أساليب أخرى يمكن بوساطتها سحب عيّنات جزئية. فإذا اختيرت الوحدة i عددًا من المرات يساوي t_i، فإن أحد البدائل هو سحب عيّنة جزئية واحـدة حجمهـا $t_i m_i$ بدون إعـادة، شريطة أن يكـون $M_i > m_i t_i$. وقـد بيّن Sukhatme (1954) أن $V(\hat{Y}_{ppz})$ ينخفض في هذه الطريقة بمقدار $(n-1)\sum^N M_i S_{2i}^2/n$.

وطريقة أخرى هي أن نسحب عيّنة جزئية واحدة حجمها m_i بصرف النظر عن عدد المـرات التي نختـار فيهـا الـوحـدة i. ويتلقى التقـدير من هذه الـوحـدة $M_i\bar{y}_i/z_i$ ترجيحة t_i (عدد المرات التي سحبت فيها الوحدة i) في أي من الطريقتين. وتأثير ذلك هو زيادة $V(\hat{Y}_{ppz})$ بمقدار،

$$\left(\frac{n-1}{n}\right)\sum^N \frac{M_i^2(1-f_{2i})S_{2i}^2}{m_i}$$

ومن أجل التكلفة نفسها، نادرًا ما تكون الفروق في الدقة بين هذه الطرق ذات بال.

وإذا كان $z_i = M_i/M_0$ فيختزل التقدير غير المنحاز في (11.31) إلى،

$$\hat{Y}_{pps} = \frac{M_0}{n}\sum_i^n \bar{y}_i \tag{11.39}$$

ومن الواضح أن هذا التقدير يصبح ذاتي الترجيح عندما يكون $m_i = m$ بحيث يصبح $\hat{Y}_{pps} = M_0\bar{\bar{y}}$.

ويكون،

$$v(\hat{Y}_{pps}) = \frac{M_0^{\,2}}{n(n-1)} \sum_i^n \left(\bar{y}_i - \frac{\hat{Y}_{pps}}{M_0}\right)^2 \qquad (11.40)$$

مقدّر عيّنة غير منحاز.

(١١-١٠) وحدات اختيرت بدون إعادة

في أي من طرق المعاينة «بدون إعادة» المدروسة في الفصل ١٩، نحصل من النظريتين (١١-١) و (١١-٢) على العلاقات الخاصة بالتباين والتباين المقدّر في معاينة على مرحلتين. وفي طريقتي Brewer أو Durbin نجد،

$$\hat{Y}_B = \sum^n \frac{M_i\bar{y}_i}{\pi_i} = \frac{1}{n}\sum^n \frac{M_i\bar{y}_i}{z_i} = \frac{1}{n}\sum^n \frac{\hat{Y}_i}{z_i} \qquad (11.41)$$

مع تباين هو،

$$V(\hat{Y}_B) = \sum_i^N \sum_{j>i}^N (\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 + \sum_i^N \frac{M_i^2(1-f_{2i})S_{2i}^{\,2}}{m_i\pi_i} \qquad (11.42)$$

ومن أجل المقدّر الموافق $\hat{Y}_{ppz}$ في معاينة مع الإعادة، نجد من (11.33) مع $\pi_i = nz_i$

$$V(\hat{Y}_{ppz}) = \frac{1}{n}\sum^N z_i\left(\frac{Y_i}{z_i} - Y\right)^2 + \sum^N \frac{M_i^2(1-f_{2i})S_{2i}^{\,2}}{m_i\pi_i} \qquad (11.43)$$

وبما أن مساهمات «ما بين الوحدات» في التباين تبقى نفسها في (11.33) و (11.43)، فإن أي مكسب نسبي في الدقة من اختيار الوحدات بدون إعادة تلغيه في عيّنات ذات مرحلتين مساهمة ما بين الوحدات في التباين، وعلى سبيل المثال، وجد Des Raj (1964) في عيّنة طبقية متعددة المراحل تتضمن $n=44$ مقاطعة من أصل $N=147$ مقاطعة، أن متوسط النسبة، $V(\hat{Y}_{WOR})/V(\hat{Y}_{WR})$ (WOR تعني بدون إعادة و WR تعني مع الإعادة)، فوق سبع مفردات كان 0.79 من أجل مركبة ما بين الوحدات، إلا إن متوسط النسبة فوق كلتا المركبتين كان 0.92 .

ومع $n=2$ يكون،

$$v(\hat{Y}_B)=(\pi_1\pi_2\pi_{12}{}^{-1}-1)\left(\frac{M_1\bar{y}_1}{\pi_1}-\frac{M_2\bar{y}_2}{\pi_2}\right)^2+\sum_{i=1}^{2}\frac{M_i^2(1-f_{2i})s_{2i}{}^2}{m_i\pi_i} \quad (11.44)$$

تقدير عيّنة غير منحاز لـ $V(\hat{Y}_B)$، حيث يرمز الدليلان 2,1 إلى الوحدات المختارة.

والتعميم إلى مقـدّر Sampford في حالـة $n>2$ لا يقـدّم أية صعـوبة. والمقدّر RHC البديل هو،

$$\hat{Y}_{RHC}=\sum_{g}^{n}\frac{Z_gM_g\bar{y}_g}{z_g} \quad (11.45)$$

حيث $Z_g=\sum z_i$ فوق الزمرة g وتشير z_g ، $\bar{y}_g$ ، وM_g إلى الوحدة المسحوبة من الزمرة. وفي عيّنة حجمها n يكون،

$$v(\hat{Y}_{RHC})=\frac{\left(\sum^{n}N_g{}^2-N\right)}{(N^2-\sum N_g{}^2)}\sum_{g}^{n}Z_g\left(\frac{M_g\bar{y}_g}{z_g}-\hat{Y}_{RHC}\right)^2+\sum_{g}^{n}Z_g\frac{M_g{}^2}{z_g}\frac{(1-f_{2g})s_{2g}{}^2}{m_g} \quad (11.46)$$

مقدّرًا غير منحاز لِـ $V(\hat{Y}_{RHC})$.

وتستـدعي العـلاقتـان (11.44) و (11.46) حسابات منفصلة لمساهمات ما بين الوحدات وما ضمن الوحدات في تقدير التباين. وفي المسوح الإحصائية المتسعة التي تتضمن العديد من الطبقات وعددًا كبيرًا من المفردات، يجعل تعقيدُ مثل هذه العلاقات تقـدير التبـاينات عملاً يحتاج من وقت الحاسب الآلي إلى أكثر مما يمكن تكريسه له عمليًا. لنتذكر الشكل الأبسط بكثير لعلاقة التباين (11.35) عندما سُحبت الوحدات مع الإعادة؛ أي العلاقة،

$$v(\hat{Y}_{ppz})=\frac{1}{n(n-1)}\sum^{n}\left(\frac{\hat{Y}_i}{z_i}-\hat{Y}_{ppz}\right)^2$$

حيث $y_i''=M_i\bar{y}_i/z_i$. وهذه النتيجة تجعل المعاينة مع الإعادة مغرية إذا لم تكن تنطوي على كثير من الخسارة في الدقة.

ومن بين طرق الـ«بـدون إعادة» يشير Platek و Singh (1972)، في تخطيطهم لإعـادة تصميم مسح القـوى العـاملة الكندية والذي تضمن $n_h=6$ أو مضاعفات

الـ 6 ، إلى مزايا البساطة وتقدير التباين التي يتمتع بها $\hat{Y}_{RHC}$ وصيغة Hartley-Rao من المعاينة النمطية (فقرة ٩ا-١٠).

وفي حالة العديد من الطبقات الصغيرة مع $n_h = 2$ أنتج Durbin (1967) طرقًا تعطي تقدير تباين بسيط عند اختيار الوحدات بدون إعادة. ويستخدم $\hat{Y}_{HT}$ متطلعًا إلى طرق يكون فيها $\pi_i = 2z_i$. وضمن طبقة معينة نهدف إلى استخدام طرق اختيار تجعل $(\pi_i \pi_j \pi_{ij}^{-1} - 1)$ إما واحدًا أو صفرًا. وتسمح هذه الاختيارات باختزال $v(\hat{Y}_{HT})$ في (11.44) إما إلى الحد،

$$\left(\frac{\hat{Y}_1}{\pi_1} - \frac{\hat{Y}_2}{\pi_2}\right)^2$$

أو إلى الحد الثاني في (11.44) وقد قام Brewer و Hanif (1969) بتحسين هذه الطرق وتوسيعها إلى مقدّرات غير المقدّر $\hat{Y}_{HT}$.

(١١-١١) مقارنة الطرق

في معاينة وحيدة المرحلة (فقرة ٩ا-١٢)، قورنت المعاينة ذات الاحتمالات المتساوية والتقدير غير المنحاز $\hat{Y}_u$ مع التقديرات النسبة المقابلة ومع طرق معاينة مختلفة باحتمالات غير متساوية. وبصورة تقريبية، كانت نسبة تباين $\hat{Y}_u$ إلى تباين طريقة أخرى عوينت بدون إعادة مساوية إلى نسبة معامل اختلاف مجموع الوحدة Y_i إلى معامل اختلاف Y_i/z_i . وفيما يتعلق بالمعاينة مع احتمالات غير متساوية بإعادة أو بدون إعادة كانت النسبة $V(\hat{Y}_{WOR})/V(\hat{Y}_{WR})$ مساوية تقريبًا لِـ $(N-n)/(N-1)$ وهو الرقم نفسه الذي نجده في حالة الاختيار باحتمالات متساوية.

وفي المعاينة متعددة المراحل وذات المرحلتين يكون التأثير الصافي لمساهمة ما ضمن الوحدات في التباين هو تخفيف الفروق الناشئة عن مساهمات ما بين الوحدات. ومع طرق المعاينة المستخدمة عمليًا، تبقى مساهمة ماضمن الوحدات النسبية الإجمالية لطرق مختلفة في اتجاه التساوي. وعمليًا لا تقل مركبة ما ضمن الوحدات في الغالب عن مركبة ما بين الوحدات.

وحتى مع الحاسبات الكبيرة تكون الأشكال ذاتية الترجيح للتقديرات مريحة ومستخدمة على نطاق واسع. وينبغي ألا يؤدي استخدام خطة ذاتية الترجيح لأكثر من خسارة طفيفة في الدقة، باعتبار أن اختيار الـ m_i لا يؤثر إلا في مساهمة ما ضمن الوحدات في التباين. وتتطلب الصيغة ذاتية الترجيح أن يكون $m_i \propto M_i/z_i$، بينما يكون الـ m_i الذي يجعل V_2 أصغر ما يمكن، من أجل عدد إجمالي متوقع مفروض للوحدات الجزئية، متناسب مع M_iS_{2i}/z_i. وتختلف هاتان المحاصتان قليلاً مالم تتغير الـ S_{2i} فوق مدى متسع.

وقد لاحظنا أيضًا في المسوح الإحصائية التي تتضمن العديد من المفردات (فقرة ١٩ـ١٣) أن الـ z_i المستخدمة في اختيار العيّنة قد لا تشكل اختيارًا جيدًا من أجل بعض المفردات، هذا عندما يكون هناك القليل من الصلة بين Y_i و z_i وإذا استطعنا إيجاد متغير عشوائي x_i بحيث تكون y_i/x_i ثابتة تقريبًا، فإن أحد الإمكانات هو أن نتحول إلى التقدير النسبة فقرة (١١ـ١٣). فقد اقترح Rao (1966) تقديرًا بديلاً من الشكل،

$$\hat{Y}^*_{ppz} = \frac{N}{n}\sum_{i}^{N} M_i\bar{y}_i \tag{11.47}$$

وتقترح بعض الدلالات التمهيدية التي توصل إليها Rao أنه إذا لم توجد صلة بين Y_i و z_i فقد يكون أداء $\hat{Y}^*_{ppz}$ في مستوى أداء $\hat{Y}_u$ نفسه، هذا إذا كانت المعاينة باحتمالات متساوية كما وضحنا في الفقرة (١٩ـ١٣). وإذا اختيرت الوحدات التمهيدية مع الإعادة يكون،

$$v(\hat{Y}^*_{ppz}) = \frac{N^2}{n(n-1)}\sum^{n} (M_i\bar{y}_i - \hat{\bar{Y}}^*_{ppz})^2 \tag{11.48}$$

تقدير عيّنة غير منحاز لِـ $V(\hat{Y}^*_{ppz})$ حيث $\hat{\bar{Y}}^*_{ppz} = \hat{Y}^*_{ppz}/N$. ولا تتضمن هذه العبارة مساهمة انحياز هذا التقدير في متوسط خطئه التربيعي، المساهمة التي نأمل أن تكون صغيرة إذا لم تكن هناك صلة بين Y_i و z_i.

(١١-١٢) النسبة إلى متغير آخر

في معاينة على مرحلتين تكون الكمية المراد تقديرها، على الغالب، نسبة Y/X. ويحدث هذا لسببين مختلفين. وكما ذكرنا سابقًا، إذا كانت x هي قيمة y في تعداد إحصائي حديث، فقد تكون النسبة y/x مستقرة نسبيًا. وقد يكون تقدير لمتوسط أو مجموع y في المجتمع، قائم على هذه النسبة، أكثر دقة من التقديرات التي استعرضناها حتى الآن في هذا الفصل.

ونواجه أيضًا التقديرات النسبة من هذا النوع، عند تقدير نسب أو متوسطات فوق أجزاء من المجتمع. وفي مسح يتم في مدينة ويعتبر الجادّة كوحدة أولية نجد كمثال على نسبة من هذا النوع،

$$\frac{\text{عدد الذكور المستخدمين ممن تزيد أعمارهم على 16 سنة}}{\text{العدد الكلي للذكور الذين تزيد أعمارهم على 16 سنة}}$$

وإذا كان $y_{ij} = 1$ لأي مستخدم ذكر فوق الـ 16 عامًا و$y_{ij} = 0$ فيما عدا ذلك، و$x_{ij} = 1$ لأي ذكر فوق الـ 14 عامًا و $x_{ij} = 0$ فيما عدا ذلك، فتكون نسبة المجتمع Y/X. وكأمثلة أخرى على هذا النوع من المسوح نذكر متوسط دخل الأُسر المشتركة في مجلة معيّنة أو متوسط مقدار العملة الموجودة في الجيب لكل طفل مراهق.

ومع أي من الطرق السابقة لاختيار الوحدات (باحتمالات متساوية أو غير متساوية، بإعادة أو بدون إعادة)، نحصل بسهولة على العلاقة التقريبية المعتادة لمتوسط مربعات خطأ أو تباين $\hat{Y}_R$ و $\hat{R}$ من العلاقة الخاصة بـ$V(\hat{Y})$ مع بقاء الطريقة نفسها في اختيار العيّنة، وذلك وفقًا للأسلوب إياه الذي استخدمناه بصورة متكررة.

وللحصول على $V(\hat{Y}_R)$ نضع $d_{ij} = y_{ij} - Rx_{ij}$ بدلاً من y_{ij} في عبارة $V(\hat{Y})$ الخاصة بطريقة المعاينة المستخدمة. ومن أجل $V(\hat{R})$ نقسم أيضًا على X^2.

ولتقدير متوسط مربعات خطأ أو تباين $v(\hat{Y}_R)$، نضع $d_{ij}' = y_{ij} - \hat{R}x_{ij}$ بدلاً من y_{ij} في عبارة $v(\hat{Y})$. وفي حالة $v_2(\hat{R})$ نقسم على $\hat{X}^2$.

وينتج هذا من الطريقة المستخدمة في النظرية (٢ ـ٥). ذلك لأن،

$$\hat{Y}_R - Y = X\frac{\hat{Y}}{\hat{X}} - Y = \frac{X}{\hat{X}}(\hat{Y} - R\hat{X}) \doteq (\hat{Y} - R\hat{X}) = \hat{D} \qquad (11.49)$$

حيث $d_{ij} = y_{ij} - Rx_{ij}$. وبما أن $E(\hat{D}) = 0$ في أي خطة معاينة يكون فيها $\hat{Y}$ و $\hat{X}$ غير منحازين، نحصل على $E(\hat{Y}_R - Y)^2$ بأخذ العلاقة الخاصة بِـ $V(\hat{Y})$ في هذه الخطة، ونعوض فيها كل y_{ij} بِـ $d_{ij} = y_{ij} - Rx_{ij}$.

وعلى سبيل المثال، لنأخذ $\hat{Y}_u = (N/n)\sum^n M_i\bar{y}_i = (N/n)\sum^n \hat{Y}_i$ في معاينة باحتمالات متساوية. فمن العلاقة (11.22) في الفقرة (١١-٧) نجد،

$$V(\hat{Y}_u) = \frac{N^2}{n}(1-f_1)\frac{\sum (Y_i - \bar{Y})^2}{N-1} + \frac{N}{n}\sum^N \frac{M_i^2(1-f_{2i})S_{2i}^2}{m_i} \qquad (11.50)$$

وبالتالي، مع معاينة باحتمالات متساوية و $\hat{Y}_R = X\hat{Y}_u/\hat{X}_u$ نجد،

$$V(\hat{Y}_R) \doteq \frac{N^2}{n}(1-f_1)\frac{\sum^N (Y_i - RX_i)^2}{N-1} + \frac{N}{n}\sum^N \frac{M_i^2(1-f_{2i})S_{d2i}^2}{m_i} \qquad (11.51)$$

حيث،

$$S_{d2i}^2 = \frac{1}{M_i - 1}\sum_{i=1}^{M_i}[(y_{ij} - Rx_{ij}) - (\bar{Y}_i - R\bar{X}_i)]^2$$

والشروط التي يُختزل تحتها $\hat{Y}_R$ إلى أحد مضاعفات نسبة مجاميع العيّنات $\Sigma\Sigma y_{ij}/\Sigma\Sigma x_{ij}$ هي دائمًا الشروط نفسهاالتي تجعل المقدّر $\hat{Y}$ المقابل ذاتي الترجيح ؛ وفي هذه الحالة يكون ثابت = $f_{2i} = m_i/M_i$.

وفي تقدير التباين يعطي وضع $d_{ij}' = y_{ij} - \hat{R}x_{ij}$ بدلاً من y_{ij} في العلاقة (11.24) الخاصة بِـ $v(\hat{Y}_u)$

$$v(\hat{Y}_R) \doteq \frac{N^2(1-f_1)}{n}\frac{\sum (\hat{Y}_i - \hat{R}\hat{X}_i)^2}{n-1} + \frac{N}{n}\sum^n \frac{M_i^2(1-f_{2i})s_{d'2i}^2}{m_i} \qquad (11.52)$$

وبصورة مماثلة، في حالة اختيار ppz مع الإعادة، وجدنا من أجل،

$$\hat{Y}_{ppz} = \frac{1}{n}\sum^n M_i\bar{y}_i/z_i$$

أن،

$$V(\hat{Y}_{ppz}) = \frac{1}{n}\sum^N z_i\left(\frac{y_i}{z_i} - Y\right)^2 + \frac{1}{n}\sum^N \frac{M_i^2(1-f_{2i})S_{2i}^2}{z_i m_i} \qquad (11.53)$$

(وذلك من (11.33) في الفقرة ١١-٩).

وبالتالي نجد من أجل $\hat{Y}_R = X\hat{Y}/\hat{X}$ في معاينة مع الإعادة،

$$V(\hat{Y}_R) \doteq \frac{1}{n}\sum^N \frac{1}{z_i}(Y_i - RX_i)^2 + \frac{1}{n}\sum^N \frac{M_i^2(1-f_{2i})S_{d2i}^2}{z_i m_i} \tag{11.54}$$

ومن (11.35) نجد تقدير العيّنة التالي $v(\hat{Y}_R)$ وهو تقدير منحاز قليلاً:

$$v(\hat{Y}_R) = \frac{1}{n(n-1)}\sum^n \left(\frac{\hat{Y}_i - \hat{R}\hat{X}_i}{z_i}\right)^2 \tag{11.55}$$

وفي طريقة Brewer في المعاينة بدون إعادة يكون $\hat{Y}_{RB} = X\hat{Y}_B/\hat{X}_B$، حيث $\hat{Y}_B = \sum M_i\bar{y}_i/\pi_i$ وفي حالة $n=2$، تعطي العلاقات (11.42) و (11.44)، الخاصة بالتقدير النسبة:

$$V(\hat{Y}_{RB}) = \sum_i^N \sum_{j>i}^N (\pi_i\pi_j - \pi_{ij})\left(\frac{D_i}{\pi_i} - \frac{D_j}{\pi_j}\right)^2 + \sum_{i=1}^N \frac{M_i^2(1-f_{2i})S_{d2i}^2}{m_i\pi_i} \tag{11.56}$$

حيث $D_i = Y_i - RX_i$ والوحدتان i و j اللتان تمّ اختيارهما،

$$v(\hat{Y}_{RB}) = (\pi_i\pi_j\pi_{ij}^{-1} - 1)\left(\frac{\hat{D}_i'}{\pi_i} - \frac{\hat{D}_j'}{\pi_j}\right)^2 + \sum_{i=1}^2 \frac{M_i^2(1-f_{2i})s_{d'2i}^2}{m_i\pi_i} \tag{11.57}$$

(١١-١٣) اختيار كسور المعاينة
وكسور المعاينة الجزئية ـ احتمالات متساوية

نوقشت المسألة أولاً من أجل تقدير النسبة إلى الحجم، وذلك عند اختيار الوحدات باحتمالات متساوية. ونفترض أن كسر المعاينة $f_2 = m_i/M_i$ ثابت، بحيث يكون التقدير متوسط العيّنة لكل وحدة جزئية.

وتتضمن أبسط دالة تكلفة ثلاثة حدود:

c_u = التكلفة المثبتة لكل وحدة أولية،

c_2 = التكلفة لكل وحدة جزئية،

c_l = تكلفة وضع قائمة لكل وحدة جزئية من وحدة مختارة.

وقد اعتبرنا الحد الثالث لأن المعاين يجب أن يضع عادة قائمة بالعناصر الموجودة في كل وحدة اختيرت، ويتحقق من عددها، كي يسحب عيّنة جزئية. وبالتالي،

$$\text{التكلفة} = c_u n + c_2 \sum^n m_i + c_l \sum^n M_i$$

وهذه العلاقة غير قابلة للاستخدام كما هي، باعتبار أن التكلفة تعتمد على المجموعة الخاصة من الوحدات التي اختيرت. وبدلاً من ذلك، لنأخذ متوسط التكلفة فوق n من الوحدات، وهو يساوي،

$$E(C) = c_u n + c_2 n\bar{m} + c_l n\bar{M} = (c_u + c_l\bar{M})n + c_2 n\bar{m} = c_1 n + c_2 n\bar{m} \quad (11.58)$$

حيث يتضمن c_l الآن متوسط تكلفة وضع قائمة لوحدة واحدة.

ونحدد n و $\bar{m} = f_2\bar{M}$ بحيث نجعل $V(\bar{y})$ أصغر ما يمكن من أجل تكلفة معطاة أو العكس. ومن (11.28) في الفقرة (١١ـ ٨)، نجد بعد القسمة على $M_0{}^2 = (\bar{M}N)^2$،

$$\text{MSE}(\bar{y}) \doteq \frac{1-f_1}{n}\frac{\sum^N M_i^2(\bar{Y}_i - \bar{Y})^2}{\bar{M}^2(N-1)} + \frac{1-f_2}{n\bar{m}}\sum^N \frac{M_i}{M_0}S_{2i}{}^2$$

لنكتب،

$$S_b{}^2 = \frac{\sum^N M_i^2(\bar{Y}_i - \bar{Y})^2}{\bar{M}^2(N-1)}$$

فهذا هو التباين المرجّح بين متوسطات الوحدات لكل عنصر. وهو مشابه للتباين $S_1{}^2$ في الفقرة (١٠ـ٣)، ويختزل إلى $S_1{}^2$ إذا كانت جميع المقادير M_i متساوية، ويمكن أيضًا كتابة،

$$S_2{}^2 = \sum^N \frac{M_i}{M_0}S_{2i}{}^2$$

وهو متوسط مرجّح لتباينات ما ضمن العيّنات، ويختزل إلى $S_2{}^2$ المذكور في الفقرة (١٠ـ٣) إذا كانت جميع المقادير M_i متساوية.

ووفقًا لهذه الرموز، وباعتبار $f_2 = \bar{m}/\bar{M}$، نجد،

$$\text{MSE}(\bar{y}) = \frac{1}{n}\left(S_b{}^2 - \frac{S_2{}^2}{\bar{M}}\right) + \frac{1}{n\bar{m}}S_2{}^2 - \frac{1}{N}S_b{}^2 \quad (11.59)$$

وبتطبيق متراجحة كوشي ـ شوارتز كالمعتاد على (11.58) و (11.59) نجد،

$$\bar{m}_{opt} \doteq \frac{S_2}{\sqrt{S_b{}^2 - S_2{}^2/\bar{M}}}\sqrt{\frac{c_1}{c_2}} \quad (11.60)$$

ويمكننا هنا تطبيق الطرق المعطاة في الفقرة (١٠ـ٨) للاستفادة من معلوماتنا حول النسبتين S_2/S_b و c_1/c_2 كمرشد لاختيار $\bar{m}_{opt}$. ويمكن تناول التقدير غير المنحاز عند سحب الوحدات باحتمالات متساوية بطريقة مماثلة .

وتقدّم الفقرة التالية تحليلاً أكثر شمولاً لهذه المسألة .

(١١-١٤) احتمالات الاختيار المثلى ومعدّلات المعاينة والمعاينة الجزئية

يحدد تحليل مهم سابق لِـ Hansen و Hurwitz (1949) في الوقت نفسه قيم z_i المثلى كدوال في الـ M_i والكسور المثلى للمعاينة والمعاينة الجزئية . ويكون اختيار الوحدات مع الإعادة . وقد قُدّم التحليل من أجل $\hat{Y}_R$ في الصيغة ذاتية الترجيح ، بحيث إن $m_i = f_0M_i/nz_i = f_0M_i/\pi_i$.

وكما في الفقرة (١١-١٣) تكون دالة التكلفة ،

$$C = c_u n + c_2 \sum^{n} m_i + c_l \sum^{n} M_i$$

ومع أن الـ M_i معروفة فالدالة تتضمن تكلفة وضع قائمة بوحدات العيّنة ، باعتبار أننا قد نحتاجها لتزويدنا بإطار للمعاينة الجزئية . ومتوسط التكلفة لمعاينة n من الوحدات هو،

$$E(C) = c_u n + C_2 f_0 M_0 + c_l \sum^{N} \pi_i M_i \qquad (11.61)$$

وبالاستناد إلى المعادلة (11.54) في الفقرة (١١-١٢) نجد،

$$V(\hat{Y}_R) = \frac{1}{n}\left[\sum^{N} \frac{1}{z_i}(Y_i - RX_i)^2 + \frac{M_i(M_i - m_i)}{z_i m_i} S_{d2i}^2\right] \qquad (11.62)$$

وبما أن $d_{ij} = y_{ij} - Rx_{ij}$ ، فيمكن كتابة $(Y_i - RX_i) = M_i\bar{D}_i$. وبملاحظة أن $\pi_i = nz_i$، $M_i/nz_i m_i = 1/f_0$، وضمّ الحدين الأول والثالث في (11.62) نجد،

$$V = V(\hat{Y}_R) = \sum^{N}\left[\frac{M_i^2}{\pi_i}\left(\bar{D}_i^2 - \frac{S_{d2i}^2}{M_i}\right) + \frac{M_i}{f_0} S_{d2i}^2\right] \qquad (11.63)$$

والمسألة هي اختيار n ، f_0 الـ $\pi_i=nz_i$ بحيث نجعل V أصغر ما يمكن، خاضعًا لمتوسط تكلفة مثبّت وإلى القيود،

$$\sum^N z_i = 1, \qquad \sum^N \pi_i = n$$

وبأخذ λ و μ كمضروبي لاغرانج وحساب القيمة الصغرى لِـ

$$V+\lambda\left[c_u n + c_2 f_0 M_0 + c_l \sum^N \pi_i M_i - E(C)\right] + \mu\left(n - \sum^N \pi_i\right) \qquad (11.64)$$

نجد بعد الاشتقاق،

$$n: \quad \lambda c_u + \mu = 0; \qquad \mu = -\lambda c_u \qquad (11.65)$$

$$\pi_i: \quad \frac{-M_i^2}{\pi_i^2}\left(\bar{D}_i^2 - \frac{S_{d2i}^2}{M_i}\right) + \lambda c_l M_i - \mu = 0 \qquad (11.66)$$

وتقود العلاقتان (11.65) و (11.66) إلى :

$$z_i = \frac{\pi_i}{n} \propto M_i\left(\bar{D}_i^2 - \frac{S_{d2i}^2}{M_i}\right)^{1/2} \Big/ (c_u + c_l M_i)^{1/2} \qquad (11.67)$$

وبما أن القيم الفردية لِـ $(\bar{D}_i^2 - S_{d2i}^2/M_i)$ سوف لا تكون معروفة، فسندرس كيف يمكن أن تعتمد القيمة المتوسطة على حجم الوحدة M_i ، مستخدمين المناقشة التقريبية التالية. لنفرض أن مجتمعا ما قد قُسّم إلى وحدات حجمها M. وبما أن $E(\bar{D}_i) = 0$ فتعطي العلاقة (9.10) من الفصل التاسع،

$$E(\bar{D}_i^2) = V(\bar{D}_i) \doteq \frac{S_d^2}{M}[1+(M-1)\rho_M]$$

حيث S_d^2 تباين المجتمع للمقادير d_{ij} و ρ_M الارتباط ضمن الوحدة لوحدات حجمها M . ولدينا أيضًا من (9.15) ،

$$E(S^2_{d2i}) \doteq S_d^2(1-\rho_M)$$

وبالتالي،

$$E\left(\bar{D}_i^2 - \frac{S_{d2i}^2}{M_i}\right) \doteq \frac{S_d^2}{M}[1+(M-1)\rho_M - (1-\rho_M)] = \rho_M S_d^2$$

ومن (11.67) يعطي هذا كتقريب،

$$z_i \propto \frac{M_i\sqrt{\rho_{M_i}}}{\sqrt{c_u + c_l M_i}} \qquad (11.68)$$

وفي حالـة ρ موجب، يمكن أن نتـوقع تناقص ρ_{M_i} مع تزايد M_i ، باعتبار أن الوحدات الجزئية البعيدة عن بعضها أقل خضوعًا للتأثيرات المشتركة، إلا أن تناقص $\sqrt{\rho_{M_i}}$ يمكن أن يكون طفيفًا. والاستنتاجات من (11.68) هي كما يلي.

١ ـ إذا كانت تكلفة وضع قائمة $c_l M_i$ غير مهمة، فإن $z_i \propto M_i$ (أي اختيـــار ا م ح) هو الأفضل شريطة أن يكون تغير $\sqrt{\rho_{M_i}}$ فوق مدى الحجوم في المجتمع تغيراً بسيطًا. أما إذا كان $\sqrt{\rho_{M_i}}$ يتناقص بصورة ملحوظة فالاحتمالات المثلى تقع بين $z_i \propto M_i$ و $z_i \propto \sqrt{M_i}$.

٢ ـ إذا كانت تكلفة وضع قائمة هي المسيطرة فينبغي أن يقع z_i بين $\sqrt{M_i}$ وعدد ثابت (احتمالات متساوية).

٣ ـ إذا كانت التكاليف المثبتة وتكاليف وضع القوائم من المرتبة نفسها في الكبر، فقد يشكل $z_i \propto \sqrt{M_i}$ حلًا وسطًا جيدًا.

واشتقاق (11.64) بالنسبة لكسر المعاينة الإجمالي f_0 يعطي،

$$f_0^2 = \frac{\sum^N M_i S_{d2i}^2}{\lambda c_2 M_0} \tag{11.69}$$

ونعثر على قيمة λ بدلالة المقادير π المعروفة، بجمع (11.66) فوق جميع الوحدات. وتقود هذه الخطوة إلى النتيجة،

$$f_0 = \left[\frac{(c_u + c_l \bar{M}) \sum^N (M_i/M_0) S_{d2i}^2}{c_2 \sum \frac{M_i^2}{\pi_i^2}\left(\bar{D}_i^2 - \frac{S_{d2i}^2}{M_i}\right)} \right]^{1/2} \tag{11.70}$$

وستبيّن المقارنـة مع (11.60) أن لـ f_0 البنية نفسهـا كما في المعـاينة باحتمالات متساوية، متذكرين أن $f_0 = n\bar{m}_{opt}/M_0$ ، $c_1 = c_u + c_l M$ في (11.60) .

ونجد القيمة المثلى لـ n من معادلة التكلفة المتوسطة (11.61) .

(١١-١٥) معاينة طبقية ـ مقدّرات غير منحازة

لا يقدّم التعميم إلى المعاينة الطبقية أية صعوبة في حالة الطرق غير المنحازة ($\hat{Y}_u$ ، $\hat{Y}_B$، $\hat{Y}_{ppz}$ الخ). ويرمز الدليل h إلى الطبقة.

وتقدير مجموع المجتمع هو $\hat{Y}_{st} = \sum \hat{Y}_h$، حيث،

$$V(\hat{Y}_{st}) = \sum_h^L V(\hat{Y}_h), \qquad v(\hat{Y}_{st}) = \sum_h^L v(\hat{Y}_h) \tag{11.71}$$

ويمكن الحصول على هذه التباينات من العلاقات المعطاة سابقًا.

ويمكننا ملاحظة الشروط التي تصبح التقديرات معها ذاتية الترجيح.

ففي حالة $\hat{Y}_{ppz}$ (فقرة ١١-٩)، أو $\hat{Y}_B$ (فقرة ١١-١٠)، نجد،

$$\hat{Y}_{st} = \sum_h^L \frac{1}{n_h} \sum_i^{n_h} \frac{M_{hi} y_{hi}}{m_{hi} z_{hi}} \tag{11.72}$$

حيث y_{hi} المجمـوع فوق الـوحـدات الجـزئية الـ m_{hi} المأخوذة من الوحدة i في الطبقة h . ويمكن رؤية أن هذه التقديرات ذاتية الترجيح ضمن الطبقات، إذا كان احتمال اختيار أي وحدة جزئية $f_{0h} = n_h z_{hi} m_{hi}/M_{hi}$ في الطبقة h ثابتًا ضمن الطبقة. وفي هذه الحالة يصبح التقدير،

$$\hat{Y}_{st} = \sum_h^L \frac{1}{f_{0h}} \sum_i^{n_h} y_{hi} \tag{11.73}$$

وهو ذاتي الترجيح تمامًا إذا بقي f_{0h} نفسه في جميع الطبقات، كما قد نتوقع بالبداهة.

وإذا كانت الوحدات بالحجم نفسه ضمن طبقة معطاة (أي $M_{hi} = M_h$)، فقد بُرهن (فقرة ١٠-١٠) أن محاصّة العيّنة التي تقود إلى تقدير ذاتي الترجيح تمامًا تكون قريبة من الـوضـع الأمثل وذلك شريطة أن يكون $S_{2h}/\sqrt{c_{2h}}$ ثابتًا تقريبًا. وعندما تتغير الـM_{hi} ضمن الطبقات، تكون النتيجة المقابلة لمقدّرات مثل $\hat{Y}_{ppz}$ و $\hat{Y}_B$ كما يلي: لنفرض أن دالة التكلفة خطية كما في الفقرتين (١١-١٣) و(١١-١٤)، وأن المقدّر ذاتي الترجيح ضمن الطبقات (أي أن $f_{0h} = n_h z_{hi} m_{hi}/M_{hi}$). فيمكن البرهان عندئذ على أن f_{0h} الموافق لتكلفة متوقعة معطاة، هو من الشكل،

$$f_{0h} \propto \frac{1}{\sqrt{c_{2h}}} \sqrt{\sum_i (M_{hi}/M_{0h}) S_{2hi}^2} \tag{11.74}$$

حيث $M_{0h} = \Sigma M_{hi}$. وهكذا فإن اختيار f_{0h} مساويًا لـ f_0 والذي يجعل هذه التقديرات ذاتية الترجيح تمامًا، سيكون، وإلى الحدّ الذي يتعلق بالدقة، قريبًا من الوضع الأمثل،

وذلك ما لم تكن المقادير S^2_{2hi} أو المقادير c_{2h} متغيرة بصورة واسعة من طبقة إلى طبقة. وهذا صحيح بصورة خاصة طالما أن هذا الاختيار يؤثر فقط في مساهمة المرحلة الثانية في التباين.

(١١-١٦) معاينة طبقية ـ المقدّرات النسبة

تتبع العلاقات الخاصة بالتقدير $\hat{Y}_{Rs}$ من تلك الموجودة في الفقرة (١١-١٢) لطبقة بمفردها، مفترضين عيّنة كبيرة في كل طبقة ومعاينة مستقلة في الطبقات المختلفة.

وفي حالة التقدير المركب، $\hat{Y}_{Rc} = X\hat{Y}_{st}/\hat{X}_{st}$ نجد،

$$\hat{Y}_{Rc} - Y = X\sum^{L} \hat{Y}_h / \sum^{L} \hat{X}_h - Y$$

$$= \frac{X}{\hat{X}_{st}} \sum^{L} (\hat{Y}_h - R\hat{X}_h) \doteq \sum^{L} (\hat{Y}_h - R\hat{X}_h)$$

وبالتالي نعوّض $d_{hij} = y_{hij} - Rx_{hij}$ بدلاً من y_{hij} لنحصل على العلاقات التقريبية لـ $V(\hat{Y}_{Rc})$ من تلك الخاصة بـ $V(\hat{Y}_{st})$ في خطة المعاينة المستخدمة. ومن أجل $v(\hat{Y}_{Rc})$ نعوّض في $v(\hat{Y}_{st})$، $d'_{hij} = y_{hij} - \hat{R}_c x_{hij}$.

وعلى سبيل المثال، في حالة احتمالات غير متساوية مع الإعادة (فقرة ١١-٩)، تقود العلاقة (11.33) إلى:

$$V(\hat{Y}_{Rc}) \doteq \sum_{h}^{L} \frac{1}{n_h} \sum_{i}^{N_h} \left[z_{hi} \left(\frac{D_{hi}}{z_{hi}} - D_h \right)^2 + \frac{M_{hi}^2 (1 - f_{2hi}) S^2_{d2hi}}{z_{hi} m_{hi}} \right] \qquad (11.75)$$

حيث،

$$D_{hi} = Y_{hi} - RX_{hi}$$

$$S^2_{d2hi} = \frac{1}{(M_{hi} - 1)} \sum_{j}^{M_{hi}} [(y_{hij} - Rx_{hij}) - (\bar{Y}_{hi} - R\bar{X}_{hi})]^2$$

ولتقدير التباين، تعطي العلاقة (11.35)، بعد التعويض،

$$v(\hat{Y}_{Rc}) \doteq \frac{\sum_{h}^{L} (D_{hi}' - \bar{D}_h')^2}{n_h (n_h - 1)} \qquad (11.76)$$

حيث،

$$D_{hi}' = \frac{M_{hi}\,\bar{d}_{hi}'}{z_{hi}}, \qquad \bar{D}_h' = \frac{\sum_i^{n_h} D_{hi}'}{n_h}, \qquad \bar{d}_{hi}' = \bar{y}_{hi} - \hat{R}_c\bar{x}_{hi}$$

(١١-١٧) مُقدّرات غير خطية في مسوح إحصائية معقدة

بالإضـافـة إلى تقـدير المجـاميع، المتوسطات، النسب، والفروق بينها، فقد ينطوي تحليل المسوح الإحصائية على تقديرات أكثر تعقيدًا من حيث البنية الرياضية (مثلًا معاملات الارتباط البسيط والجزئي، الوسط، ومئينات أخرى). وقد تتضمن أهداف التحليل وضع فترة ثقة للكمية المقدّرة أو القيام باختبار أهمية.

لنفرض أن لدينا عيّنة عشوائية من مجتمع لانهائي، فقد أنتج الإحصاء النظري طرقًا متنوعة تواجه هذه الأهداف ـ طرق مضبوطة للعيّنات الصغيرة قائمة على فروض تتعلق بطبيعـة التـوزيعـات التي تتبعها الملاحظات، وطرق تقريبية للعيّنات الكبيرة تستدعي عددًا أقل من الفروض. وعلى الوجه الآخر، كنا قادرين، من أجل أنواع العيّنة الأكثر تعقيدًا والمدروسة في هذا الكتاب، على إعطاء طرق لحساب تقديرات غير منحازة لتباينات تقديرات خطية غير منحازة مثل $\hat{Y}_{ppz}$ $\hat{Y}_{st}$ أو $\hat{Y}_{HT}$. ومفترضين أن العيّنة كبيرة بما يكفي للقول بأن لهذه التقديرات توزيعات قريبة من التوزيع الطبيعي، فإن جداول التـوزيع الطبيعي تقدّم حدود ثقة واختبارات أهمية. ويبقى العديد من المسائل المتعلقة بمقدّرات غير خطية في مسوح إحصائية معقدة.

وقد أُنتجت ثلاث طرق تقريبية لتقدير الأخطاء المعيارية لمقدّرات غير خطية. وسنعرضها في حالة عيّنات عشوائية طبقية مع $n_h=2$ في جميع الطبقات، وهي الحالة التي نالت معظم الدراسة. وجميع هذه الطرق تعطي كالعادة تقديرًا غير منحاز للتباين عندما يكون المقدّر خطيًا. ويمكن للعيّنة أن تكون متعددة المراحل حيث تسحب الوحدات الأولية إما باحتمالات متساوية أو باحتمالات غير متساوية مع الإعادة.

(١١ ـ ١٨) النشر وفق متسلسلة تايلور

هذه هي الطريقة التي أنتجت العلاقات التقريبية لتقديرات تباين $\hat{R}$ ، $\hat{R}_c$ و $\hat{R}_s$ في معاينة طبقية. ويُعبّر عن الدالة المراد تقديرها $f(Y_1, Y_2, \ldots, Y_k) = f(\mathbf{Y})$ كدالة في المجاميع فوق المجتمع لمتغيرات معيّنة، والتقدير $f(\hat{\mathbf{Y}})$ هو الدالة الموافقة في تقديرات عيّنة غير منحازة لِـ $\hat{\mathbf{Y}}$. ومن أجل متغير j مع $n_h = 2$ وعند اختيار الوحدات بطريقة أم مح مع الإعادة، يكون مجموع المجتمع وتقديره الخطي ،

$$Y_j = \sum_h^L \sum_i^{N_h} Y_{jhi}; \qquad \hat{Y}_j = \sum_h^L \sum_i^2 \frac{\hat{Y}_{jhi}}{2z_{hi}} = \sum_h^L (y'_{jh1} + y'_{jh2}) \tag{11.77}$$

حيث $y'_{jhi} = \hat{Y}_{jhi}/2z_{hi}$ و $\hat{Y}_{jhi}$ تقدير عيّنة غير منحاز لِـ Y_{jhi} محسوب من العيّنة الجزئية في هذه الوحدة.

ومن (11.35) نستنبط تقديرًا غير منحاز لِـ $V(\hat{Y}_j)$ على الشكل ،

$$v(\hat{Y}_j) = \sum_h \sum_i^2 2(y'_{jhi} - \bar{y}'_{jhi})^2 \equiv \sum_h (y'_{jh1} - y'_{jh2})^2 \tag{11.78}$$

وبالتالي، واستنادًا إلى تعميم Woodruff (1971) في الفقرة (٦-١٣) لطريقة Keyfitz المختصرة في حالة $n_h = 2$، يكون تقريب متسلسلة تايلور لتباين $f(\hat{\mathbf{Y}})$ هو،

$$v[f(\hat{\mathbf{Y}})] \doteq \sum_h \left[\sum_j^k \left(\frac{\partial f}{\partial \hat{Y}_j}\right)(y'_{jh1} - y'_{jh2})\right]^2 \tag{11.79}$$

وقد تتطلب عبارة دالة غير خطية في المتغيرات المقيسة من الشكل $f(\mathbf{Y})$ بعض الجهد والحذر، وقد لا تكون ممكنة بالنسبة لبعض الدوال ذات الأهمية. لنأخذ مثالاً بسيطًا ـ معاينة عشوائية بسيطة مع وحدات أولية ذات حجوم متساوية، حيث يراد تقدير الارتباط بين مجاميع الوحدات لمتغيرين U_1 و U_2. فقيمة المجتمع $f(\mathbf{Y})$ هي ،

$$\rho = \frac{\sum_h^L \sum_i^N U_{1hi}U_{2hi} - \dfrac{\left(\sum_h \sum_i U_{1hi}\right)\left(\sum_h \sum_i U_{2hi}\right)}{N}}{\left[\sum_h \sum_i U_{1hi}^2 - \dfrac{\left(\sum_h \sum_i U_{1hi}\right)^2}{N}\right]^{1/2} \left[\sum_h \sum_i U_{2hi}^2 - \dfrac{\left(\sum_h \sum_i U_{2hi}\right)^2}{N}\right]^{1/2}}$$

وبدلالة المتغيرات Y_{jhi} تشكل هذه العبارة دالة في خمسة متغيرات هي :

$$Y_{1hi} = U_{1h}\text{[illegible]}_{hi}; \qquad Y_{2hi} = U_{1hi}^2; \qquad Y_{3hi} = U_{2hi}^2$$

$$Y_{4hi} = U_{1hi}; \qquad Y_{5hi} = U_{2hi}$$

حيث $y'_{1hi} = \hat{U}_{1hi}\hat{U}_{2hi}/2z_{hi},\ i = 1, 2,\ z_{hi} = 1/N_h$ في هذه الحالة.

(١١-١٩) إعادة تكرارات متوازنة

نستعــرض هذه الــطريقــة أولاً في حالــة دالــة خطية في متغير بمفـرده. ليكن $f(Y) = Y$، وتقديره $f(\hat{Y}) = \hat{Y} = \sum_h (y_{h1}' + y_{h2}')$ حيث $y_{hi}' = \hat{Y}_{hi}/2z_{hi}$. فمن (11.78) نجد تقديرًا غير منحاز للتباين $v(\hat{Y})$ على الشكل،

$$v(\hat{Y}) = \sum_h (y_{h1}' - y_{h2}')^2 \tag{11.80}$$

لنختر نصف عيّنة H باختيار وحدة من كل طبقة. وتقدير $f(Y) = Y$ من نصف العيّنة هذا هو $f(H) = 2\sum_h y_{hi}'$ حيث hi الوحدة التي اخترناها من الطبقة h . وبالتالي، إذا كان $f(S)$ يرمز لتقدير $f(Y)$ المأخوذ من العيّنة بكاملها، فعندئذ،

$$f(H) - f(S) = 2\sum_h y_{hi}' - \sum_h (y_{h1}' + y_{h2}') = \sum_h \pm (y_{h1}' - y_{h2}') \tag{11.81}$$

وتعتمد الإشارات على ما إذا كانت الوحدة 1 أو الوحدة 2 قد اختيرت من الطبقة h .

وتبيّن مقارنـة (11.80) ، (11.81) أنه في أي من الاختيارات الـ 2^L الممكنة، يتضمن $[f(H) - f(S)]^2$ الحدود التربيعية الصحيحة $(y_{h1}' - y_{h2}')^2$ في $v(\hat{Y})$ وذلك في كل طبقة. وعلى أي حال، ففي أي نصف عيّنة محدد يسهم كل زوج من الطبقات بحدّ جدائي في $[f(H) - f(S)]^2$ إشـارته ± . وقد لاحظ McCarthy (1966) أنه يمكن إيجاد مجموعة متوازنة من أنصاف العيّنات تحقق ما يلي: في كل زوج من الطبقات، تكون إشارة نصف الحدود الجدائية في المجموعة موجبة وإشارة النصف الآخر سالبة. وإذا احتوت هذه المجموعة g من أنصاف العيّنات المختلفة، فإن الحدود الجدائية التي لا نريدها تلغي بعضها بعضًا عند الجمع فوق المجموعة مما يعطي،

$$\frac{1}{g}\sum_i^g [f(H_i) - f(S)]^2 \equiv \sum_h^L (y_{h1}' - y_{h2}')^2 = v(\hat{Y}) \tag{11.82}$$

وإذا كان C_i نصف العيّنة المتمم لِـ H_i ، نجد شكلاً ثانيًا وثالثًا هما ،

$$v(\hat{Y}) \equiv \frac{1}{g}\sum_i^g [f(C_i)-f(S)]^2 \equiv \frac{1}{4g}\sum_i^g [f(H_i)-f(C_i)]^2 \qquad (11.83)$$

والمقدّر الرابع هو متوسط المقدّرين الأولين .

وكان Placket و Burman (1946) قد سبقا إلى إعطاء هذا النوع من المجموعات المتـوازنة في تصميم التجارب ، وذلك من أجل g يساوي أيًا من مضاعفات الـ 4 . ومـع L من الـطبقـات تكون g في أصغر مجموعة متوازنة مساوية لأصغر مضاعفات الـ 4 التي لا تقل عن L . وفي حالة $L=5$ طبقات يكون $g=8$. ويبيّن الجدول (١١-٧) مجموعة متوازنة لحالة $L=5$ ، وترمز + للوحدة 1 كما ترمز - للوحدة 2 في طبقة .

جدول (١١-٧) أنصاف متوازنة للعيّنة في حالة L=5 طبقات

طبقة	نصف - عيّنة 1	2	3	4	5	6	7	8
1	+	+	+	−	+	−	−	−
2	−	+	+	+	−	+	−	−
3	−	−	+	+	+	−	+	−
4	+	−	−	+	+	+	−	−
5	−	+	−	−	+	+	+	−

وعنــد تطبيق طريقـــة التكــرار المعـاد المتـوازن هذه (ت ع م) على دالـة غير خطية $f(\mathbf{Y})$ في عدة متغيرات ، فالمقدّرات

$$\frac{1}{g}\sum_i^g [f(H_i)-f(S)]^2; \qquad \frac{1}{g}\sum_i^g [f(C_i)-f(S)]^2; \qquad \frac{1}{4g}\sum_i^g [f(H_i)-f(C_i)]^2$$

بالإضــافــة إلى متــوسط الأول والثــاني، تختلف جميعهــا إلى حد ما . ويكــون $f(S)=f(\hat{\mathbf{Y}})$ منحـازًا أيضًا ، ولا تتضمن تقديرات التباين مساهمة الانحياز الصحيحة في $\text{MSE}[f(\hat{\mathbf{Y}})]$. وعـلى أي حال ، نلاحظ أنـه إذا سحبنا m من العيّنات المستقلة S_i ، كل منها بِـ H_i, C_i المتتامين ، فستكون الكمية ،

$$\frac{1}{2m}\sum_i^m [f(H_i)-f(C_i)]^2 \qquad (11.84)$$

تقديرًا غير منحاز لـ $V[f(H)]$. وهكذا إذا تكلمنا بصورة تقريبية، تؤدي طريقة الـ(ت ع م) إلى أن نفترض من أجل $f(S)$ أن، (ا) انحيازه مهمل، (ب) $V[f(H)]^{(1/2)}$ $\doteq V[f(S)]$ و (ج) حساب $v[f(S)]$ من ت ع م، من أجل عيّنة طبقية واحدة S، يتفق مع حسابه من عيّنات مستقلة S_i اتفاقًا هو من الجودة بما يكفي لاعتباره حسابًا مفيدًا.

وقد استُخدمت طريقة التكرارات المعادة على نطاق واسع في كل من اختيار عيّنة وتقدير تباين من قِبل مكتب الإحصاء في الولايات المتحدة في الخمسينات كما استخدمه Deming (1960, 1956) . وتعود الفكرة إلى أسلوب Mahalanobis, في العيّنات الجزئية المتداخلة [Mahalanobis , 1944] . وقد أعاد McCarthy (1969) النظر في خواص طريقة الـ BRR مع مزيد من المناقشة من قبل Kish و Frankel (1974) .

(١١-٢٠) طريقة مدية الجيب

وُصفت هذه الطريقة في الفقرة (٦-١٧) من أجل $\hat{R}$ في عيّنات عشوائية بسيطة. وقد اقتُرحت كوسيلة لتقدير $V(\hat{R}_Q)$، ولكن يمكن أيضًا تجربتها لتقدير $V(\hat{R})$. لنحذف الوحدة j من العيّنة ونحسب التقدير النسبة $\hat{R}_j$ ، من بقية العيّنة. ومتجاهلين الـ ت م م يمكن الحصول على مقدّر تقريبي لـ $V(\hat{R})$ من (6.86) مع $g=n$. وإذا افترضنا $\bar{\hat{R}} \doteq \hat{R}$ أي متوسط الـ $\hat{R}_j$ يصبح هذا المقدّر،

$$v(\hat{R}) \doteq \frac{(n-1)}{n}\sum_i^n(\hat{R}_j-\hat{R})^2 = \frac{(n-1)}{n}[f(S_j)-f(S)]^2 \qquad (11.85)$$

وتشير العبارة في أقصى اليمين إلى الكيفية التي يمكن فيها تطبيق الطريقة على تقدير آخر غير خطي.

وفي حالة مقدّر خطي مثل $\bar{y}$ يسهل التحقق من أن العلاقة (11.85) تُختزل إلى التباين المعتاد $v(\bar{y}) = \sum(y_i-\bar{y})^2/n(n-1)$.

وفي مدّ هذه الطريقة إلى عيّنات طبقية مع $n_h=2$ يقترح Frankel (1971) حذف وحدة واحدة من الطبقة h ، بصورة عشوائية، وذلك من أجل $h = 1, 2, \ldots L$، كل بدورها، حاسبين $f(S_h)$ من العيّنة الباقية وحجمها $(2L-1)$. ويكون أحد أشكال تقدير مدية الجيب لـ $V[f(S)]$ عندئذ هو،

$$v[f(S)]=\sum_{h}^{L}[f(S_h)-f(S)]^2 \qquad (11.86)$$

وفي حالة مقدّر خطي $f(\hat{Y})$، يُختزل مقدّر التباين إلى المقدّر غير المنحاز المعتاد في معاينة ا م ح مع الإعادة.

وكما في ت ع م*، توجد أربع نسخ متشابهة لمقدّر مدية الجيب لـ $v[f(S)]$.

(١١-٢١) مقارنة الطرق الثلاث

تمتلك طرق Taylor وت ع م، ومـدية الجيب قدرًا محدودًا جدًّا من التـبرير النـظري الدقيق. وتثمين إنجازاتها، مع مقدّرات مختلفة، وأنواع مختلفة من المسوح الإحصـائية، قد اعتمـد حتى الآن اعتـمادًا رئيسًـا على دراسـات مونت كارلو. وقد وصف Frankel (1971) و Kish, Frankel (1974) إحدى الدراسات.

كانت العيّنـة عيّنـة وحيدة المرحلة تتضمن وحدات عنقودية تختلف حجومها اختـلافًـا طفيفًا، وتتضمن الوحدة في المتوسط 14.1 أسرة. وقد قُسم المجتمع الذي يتضمن N=3240 وحـدة، أو 45737 أسرة، إلى 12,6 ، و 30 من الـطبقات المتساوية الحجم، على التـوالي، مع n_h=2 في كل طبقة. واستُخدمت معاينة عشوائية بسيطة ضمن الـطبقـات. وقـد أُخـذت البيانـات الإحصـائية من الـ Population Survey (U.S Census Bureau) Current . وكانت المتغيرات الثمانية الأصلية من أجل أسرة هي عدد الأفـراد، العدد تحت سن الـ 18 ، عدد الأفراد ضمن القوة العاملة، الدخل، العمر، الجنس، وعدد سنوات المدرسة لربّ الأسرة، والدخل الإجمالي. وقد حُسبت متوسطات (نسب) مختلفة، فروق بين متوسطات، ومعاملات انحدار أو ارتباط، بسيط أو جزئي أو متعـدد. وفي أعـداد الـوحـدات الأولية كانت العيّنـات صغـيرة تمامًا، 24, 12 و 60 أي حوالي 170, 340 و 850 بالنسبة لأعداد الأُسر.

وبـالنسبـة لحسـاب مقـدّرات التبـاين قورنت جميع الأشكـال الأربعة لطرق الـت ع م ومـدية الجيب. ولم تُستخـدم طريقـة Taylor من أجل الانحدار المتعدد

* ت ع م اختصار لتكرار معاد متوازن.

الجزئي، بسبب صعوبة التعبير عن المشتقات الجزئية $\partial f/\partial Y_j$ في شكل طيّع قابل للاستخدام. وبما أن طريقة المعاينة كانت طبقية تناسبية بدون إعادة، فقد تضمنت جميع تقديرات التباين الـ ت م م وهو $(1-f)$، حيث $f=n/N$، مع أننا نادرًا ما نحتاجها مع مثل هذه العيّنات الصغيرة.

وباستثناء ما يتعلق بمعاملات الارتباط المتعدد، كان انحياز المقدّرات $f(\mathbf{Y})$ غير مهم نسبيًا، وذلك من أجل الحجوم الثلاثة للعيّنة جميعها، وكانت النسب $\frac{\text{الانحياز}}{\text{الخطأ المعياري}}$ أقل من 0.1 من أجل نسب، فروق بين نسب، ومعاملات انحدار بسيط، وكانت حوالي 0.2 من أجل معاملات ارتباط بسيط ومتعدد. وعندما اعتُبرت قيمة $v(\hat{\mathbf{Y}})$ التقريبية كتقديرات لِـ $\text{MSE}(\hat{\mathbf{Y}})$ كانت إنجازات الطرق الثلاث جميعها BRR Taylor (ت ع م)، ومدية الجيب-جيدة من أجل نسب وفروق بين نسب، حيث كان الانحياز في $v(\hat{\mathbf{Y}})$ مقسومًا على $\text{MSE}(\hat{\mathbf{Y}})$ تحت الـ 5% . ومن أجل معاملات ارتباط بسيطة كان لمقدّرات التباين في طريقة الـت ع م انحيازات أصغر بكثير من مقدّرات طريقة Taylor (T) وطريقة مدية الجيب (J) . وكان العكس هو الصحيح من أجل معاملات الانحدار البسيط. ومع كل من BRR ومدية الجيب كان متوسط التقديرين H و C (أي مع ت ع م) أفضل الطرق الأربعة لِـ $v(\hat{Y})$.

$$\frac{1}{2g}\left\{\sum_{i}^{g}[f(H_i)-f(S)]^2+\sum_{i}^{g}[f(C_i)-f(S)]^2\right\} \tag{11.87}$$

وفي عبارات فترات الثقة واختبارات الأهمية، تكون المسألة المهمة هي إلى أيّ حد يتفق احتمالا الذيلين في المتغير $[f(\hat{\mathbf{Y}})-f(\mathbf{Y})]/\text{s.e.}[f(\hat{\mathbf{Y}})]$ مع ذيلي التوزيع t بـ L درجة من الحرية $s.e.[f(\hat{\mathbf{Y}})]$ يعني الخطأ المعياري لِـ $f(\hat{\mathbf{Y}})$ ومن أجل قيم مختارة لِـ t يعطي Frankel (1971) تقديرات مونت كارلو لتكراري الذيلين من أجل المتغير $[f(\hat{\mathbf{Y}})-Ef(\hat{\mathbf{Y}})]/\text{s.e.}[f(\hat{\mathbf{Y}})]$ ويبيّن الجدول (١١ ـ ٨) بعض النتائج في حالة 6 طبقات و30 طبقة، وذلك من أجل قيم t التي يقدّمها Frankel وهي أقرب ما يمكن لمستويي 5 و 10% الثنائيي الذيل. والنسخة (11.87) لطريقتي، ت ع م، ومدية الجيب (J) هي النسخة المبيّنة في الجدول.

جدول (١١-٨) متوسط الاحتمالات الذيلية لـ $[f(\hat{Y}) - Ef(\hat{Y})]/\text{s.e.}[f(\hat{Y})]$ مقارنة مع تلك الخاصة بتوزيع ستيودنت t

	ست طبقات					
	$P(t) = .042, t = 2.576$			$P(t) = .098, t = 1.960$		
	BRR	J	Taylor	BRR	J	Taylor
نسب	.044	.049	.052	.096	.106	.112
b's[a]	.034	.048	.058	.085	.117	.127
r's[b]	.052	.069	.084	.114	.137	.163
r جزئي	.043	.063	—	.092	.132	—
R متعدد	.065	.088	—	.105	.160	—
	30 طبقة					
	$P(t) = .059, t = 1.960$			$P(t) = .110, t = 1.645$		
	BRR	J	Taylor	BRR	J	Taylor
نسب	.056	.057	.057	.109	.111	.112
b's	.062	.067	.068	.110	.116	.116
r's	.089	.098	.102	.138	.153	.164
r جزئي	.103	.121	—	.156	.181	—
R متعدد	.175	.207	—	.265	.297	—

معامل انحدار بسيط = b[a]

معامل ارتباط بسيط = r[b]

ونلخص بقولنا أن ت ع م تنجز بثبات إنجازًا أفضل في الجدول (١١-٨). وباستثناء ما يتعلق بالمقادير «R المتعدد»، يمكن اعتبار هذه الطريقة مناسبة للاستخدام العملي، إذا توافرت لدينا القناعة أنه عند تحليل البيان الإحصائي تمثل قيمة ذيلية جدولية مقدارها 5% قيمةً ذيلية فعلية في مكان ما بين 3% و 8% . وطريقة مدية الجيب (J) أفضل بقليل من Taylor وباستثناء ما يتعلق بـ ت ع م في حالة نسب وانحدارات بسيطة، تعطي كل الطرق تكرارت ذيلية فعليه أعلى مما في جداول t ، بحيث تكون احتمالات الثقة مبالغًا فيها. والأمر المحيّر هو أنه من أجل معاملات الارتباط لا تؤدي الزيادة في حجم العيِّنة من 12 إلى 60 إلى تحسين مقابل في اقتراب التكرارات الذيلية الفعلية من تكرارات t الذيلية.

وتفتح هذه الدراسة ميدانًا واسعًا لتقصيّ طرق بخطط مختلفة للمسح وبأنواع مختلفة من المقدرات $f(\mathbf{Y})$.

وفي دراسة مونت كارلو لعيّنة أكبر وأكثر تعقيدًا (معاينة أ م ح على مرحلتين مع الإعـادة، متـضـمنـة تقسيمًا إلى طبقـات وتقسيمًا لاحقًـا إلى طبقـات) قارن Bean (1975) طريقتي Taylor وت ع م من أجـل تقديرات من نوع التقدير النسبة. وأعطت كلتا الطريقتين، تقديرات تباين مُرضية، واحتمالات ثقة ثنائية الجانب مناسبة، محسوبة من التوزيع الطبيعي، وعلى أي حال، فقد بقي التواء كاف بحيث لا يمكن الاطمئنان إلى فترات الثقة وحيدة الجانب.

تمارين

(١١-١) احسب التقديرات من أجل جميع العيّنات الممكنة التي يمكن سحبها من المجتمع الاصطناعي في الجدول (١١-١)، وذلك بالطرق II,*Ib*,I*a* و III، تحقّق من قيم الـ MSE (متوسط الخطأ التربيعي) الكلي المعطى في الجدول (١١-٢).

(١١-٢) من أجـل الـطرق II (احتـمالات متسـاوية، تقـدير غير منحـاز)، و III (اختيار ا م ح)، أعـد حسـاب تبـاينـات $\hat{Y}$ في مثـال الجدول عندما m_i=1 . بيّن أن دقة الطريقة III بالنسبة للطريقة II أقل في حالة m_i=1 مما هي في حالة m_i=2 . ما هي النتيجة العامة التي يوضحها هذا ؟

(١١-٣) من أجـل المجتمع في الجدول (١١-١)، إذا كانت الحجوم المقدّرة z_i هي 0.1 ، 0.3 و 0.6 ، مع m_i=2 بيّن أن التقدير غير المنحاز (طريقة IV) يعطي تباينًا أصغر من تباين المعاينة ا م ح. ما هو تفسير هذه النتيجة؟

(١١-٤) صُنفت العناصر في مجتمع يتضمن ثلاث وحدات أولية إلى صفّين، وحجوم الوحدات M_i والنسب P_i للعناصر التي تنتمي إلى الصف الأول كانت كما يلي:

$$M_1 = 100, \quad M_2 = 200, \quad M_3 = 300, \qquad P_1 = 0.40, \quad P_2 = 0.45, \quad P_3 = 0.35$$

في عيّـنة تتـألف من 50 عنصرًا من إحدى الوحدات الأولية، قارن متوسطات الخطأ التربيعية للطرق I_a ، II وIII الخاصة بتقدير نسب عناصر الصف الأول في المجتمع. (في علاقات التباين في الفقرة (١١-٢)، s_i^2 هي تقريبًا (P_iQ_i)).

(١١-٥) اختيرت عيّنة من n من الوحدات الأولية باحتمالات متساوية. ومن كل وحدة مختارة، نأخذ كسرًا ثابتًا f_2 من الوحدات الجزئية. وإذا وقعت a_i من الـ m_i وحدة

جزئية في الوحدة i في الفصل C ، فبيّن أن تقدير النسبة إلى الحجم (فقرة ١١-٨) لنسبة عناصر المجتمع هو $\bar{p}=\Sigma a_i/\Sigma m_i$ بيّن من العلاقة (11.36) أن،

$$v(\bar{p})=\frac{1-f_1}{n\bar{M}^2}\frac{\sum^n M_i^2(p_i-\bar{p})^2}{n-1}+\frac{f_1(1-f_2)}{n^2\bar{m}\bar{M}}\sum^n\frac{M_im_i}{m_1-1}p_iq_i$$

هو تقدير لِـ $\text{MSE}(p)$ حيث $p_i=a_i/m_i$.

(١١-٦) تقرر شركة تمتلك 36 مصنعًا، التحقق من شروط أحد التجهيزات التي تتضمن $M_0=25{,}012$ قطعة قيد الاستخدام. أُخذت عيّنة عشوائية من 12 مصنعًا، ودُققت عيّنة جزئية نسبتها 10% من كل مصنع جرى اختياره. وكانت أعداد القطع المدققة (m_i) وأعداد القطع التي عُثر فيها على بوادر فساد (a_i) كما يلي:

المصنع	m_i	a_i	$p_i=\frac{a_i}{m_i}$	المصنع	m_i	a_i	$p_i=\frac{a_i}{m_i}$
1	65	8	0.123	7	85	18	0.212
2	82	21	0.256	8	73	11	0.151
3	52	4	0.077	9	50	7	0.140
4	91	12	0.132	10	76	9	0.118
5	62	1	0.016	11	64	20	0.312
6	69	3	0.043	12	50	2	0.040

قدّر النسبة المئوية والعدد الكلي للقطع الناقصة الصنع الموجودة قيد الاستخدام، واعط تقديرات لأخطائها المعيارية.

ملاحظة: بما أن $M_i/\bar{M}\doteq m_i/\bar{m}$ فيمكن حساب مركبة ما بين الوحدات في $v(\bar{p})$ على الشكل،

$$\frac{1-f_1}{n\bar{m}^2(n-1)}\left(\sum a_i^2-2\bar{p}\sum a_im_i+\bar{p}^2\sum m_i^2\right)$$

وبما أن الـ m_i كبيرة نسبيًا، فيمكن أيضًا حساب مركبة ما ضمن الوحدات على الشكل،

$$\frac{f_1(1-f_2)}{(n\bar{m})^2}\sum a_iq_i$$

(١١-٧) إذا اختيرت الوحدات الأولية باحتمالات متساوية وكان f_2 ثابتًا فبيّن أنه

وفق رموز التمرين (١١-٥)، يكون التقدير غير المنحاز لنسبة مجتمع هو $p = N\Sigma a_i / nM_0 f_2$ وأنه إذا كانت الـ $1/m_i$ مهملة، فيمكن حساب تباينه على الشكل،

$$v(p) = \frac{1-f_1}{n(n-1)\bar{m}^2} \sum^{n} (a_i - \bar{a})^2 + \frac{f_1(1-f_2)}{(n\bar{m})^2} \sum^{n} a_i q_i$$

احسب p وخطأه المعياري من البيان الإحصائي في التمرين (١١-٦).

(١١-٨) اختيرت عيّنة من n من الوحدات الأولية باحتمالات ثنائية مع تقديرات الحجم z_i (مع الإعادة) وكسر معاينة إجمالي (متوقع) ثابت f_0. بيّن أن التقدير غير المنحاز، وتقدير النسبة إلى الحجم، لمجموع المجتمع هما، على الترتيب، T/f_0 و $TM_0/\sum^{n} m_i$ حيث T مجموع العيّنة. (ونستنتج أنه إذا كان M_0 غير معروف فيمكن استخدام التقدير غير المنحاز دون تقدير النسبة إلى الحجم. وتنعكس الحالة عند تقدير متوسط المجتمع لكل وحدة جزئية).

(١١-٩) في دراسة للازدحام في مدينة كبيرة تضمنت طبقة 100 جادة اختير عشر منها باحتمالات متناسبة مع تقدير الحجم (مع الإعادة). واستُخدم كسر معاينة إجمالي (متوقع) 2% . قدّر العدد الكلي للأشخاص ومتوسط عدد الأشخاص لكل غرفة وأخطائها المعيارية وذلك من البيان الإحصائي أدناه،

جادّة	1	2	3	4	5	6	7	8	9	10
غرف	60	52	58	56	62	51	72	48	71	58
أشخاص	115	80	82	93	105	109	130	93	109	95

(١١-١٠) في طريقة Durbin (فقرة ١١-١٠) لتبسيط تقدير التباين مع معاينة ا م ح بدون إعادة، نعرض فيما يلي طريقة بسيطة لاختيار العيّنة، تعود أساسًا إلى Kish (1965) . سنلغي الدليل h الذي يرمز للطبقة، ونفرض أن عدد الوحدات الأولية زوجي.

نرتب الوحدات وفق قيم z_i المتصاعدة ثم نؤشرها زوجًا زوجًا. وتكون الطريقة مضبوطة فقط إذا كان $z_i = z_j$ من أجل وحدتين في الزوج نفسه؛ وسنفترض أن هذا محقق هنا. نختار وحدتين ام ح (ppz) الإعادة. إذا سحبنا وحدتين مختلفتين فنقبلهما معًا. وإذا سُحبت الوحدة نفسها مرتين فتتألف العيّنة من وحدتي الزوج الذي تنتمي إليه هذه

الوحدة. بيّن أنه من أجل هذه الطريقة: (ا) $\pi_i = 2z_i$، (ب) في حالة وحدات ليست في الزوج نفسه $\pi_{ij} = 2z_i z_j = \pi_i \pi_j / 2$، بحيث يكون $\pi_i \pi_j \pi_{ij}^{-1} - 1 = 1$. و(ج) في حالة وحدات في الزوج نفسه $\pi_{ij} = 4z_i z_j = \pi_i \pi_j$، بحيث يكون $\pi_i \pi_j \pi_{ij}^{-1} - 1 = 0$.

(١١-١١) في الفقـرة (١١-٩) بُرهنت العـلاقة (11.33) من أجل $V(\hat{Y}_{ppz})$ في معاينة مع الإعادة، وفق خطة تقضي أنه حيثما نختار الوحدة i، نسحب عيّنة جزئية عشوائية بسيطة حجمها m_i من الوحدة بكاملها، برهن النتائج التالية من أجل خطتين بديلتين.

(ا) عنـد اختيار الوحدة i عددًا من المرات t_i نسحب منها عيّنة جزئية عشوائية بسـيطة حجمهـا $m_i t_i$ (نفـترض $m_i t_i \leq M_i$). وتحت هذه الخـطة ينخفض $V(\hat{Y}_{ppz})$ في (11.41) بمقدار، $(n-1)\sum^{N} M_i S_{2i}^2 / n$ (1954, Sukhatme).

(ب) عند اختيار الوحدة i عددًا من المرات t_i نسحب عيّنة جزئية عشوائية بسيطة حجمها m_i. وعندئذ يزداد $V(\hat{Y}_{ppz})$ في (11.41) بمقدار.

$$\frac{(n-1)}{n} \sum^{N} M_i^2 (1 - f_{2i}) S_{2i}^2 / m_i$$

وفي كلتا الخطتين (ا) و(ب) تتلقى الوحدة i الترجيحة $\hat{Y}_{ppz} = \sum^{N} t_i M_i \bar{y}_i / n z_i$



الفصل الثاني عشر

المعاينة المضاعفة

(١٢-١) وصف الطريقة

يعتمد عدد من طرق المعاينة، كما رأينا، على امتلاك معلومات مسبقة حول متغير مساعد x_i. ويتطلب التقدير النسبة وتقدير الانحدار معرفة بمتوسط المجتمع $\bar{X}$. وإذا كان من المرغوب تقسيم المجتمع إلى طبقات وفقًا لقيم x_i ، فيجب أن يكون التوزيع الاحتمالي لـ x_i معروفًا.

وعندما تنقصنا مثل هذه المعلومات، يكون من غير المكلف أحيانًا أن نأخذ عيّنة تمهيدية نقيس فيها x_i فقط. والغاية من هذه العَينة هي أن تمدّنا بتقدير جيّد لـ $\bar{X}$ ولتوزيع x_i. وفي مسح إحصائي وظيفته هي القيام بتقديرات تتعلق بمتغير آخر y_i ، قد يكون مجزيًا أن نكرّس لهذه العيّنة التمهيدية جزءًا من الموارد المتوافرة، مع أن هذا يعني ضرورة تخفيض حجم العيّنة في المسح الإحصائي الرئيس المتعلق بـ x_i . وتُعرف هذه الطريقة بالمعاينة المضاعفة أو المعاينة ثنائية الطور. وكما تقتضي المناقشة ضمنًا، فإن هذه الطريقة ستكون مُربحة فقط إذا كان الكسب في الدقة من التقدير النسبة، أو تقدير الانحدار، أو الكسب من التقسيم إلى طبقات، أكثر من أن يعوّض الخسارة في الدقة العائدة إلى تخفيض حجم العيّنة الرئيسة.

وقد تكون المعاينة المضاعفة ملائمة جدًا عندما تكون المعلومات حول x_i موجودة في بطاقات مصنفة لم تجر جدولتها بعد. فمثلاً في المسوح الإحصائية المتعلقة بالمجتمع السكاني المدني في ألمانيا في عام 1945 ، كانت تُسحب العيّنة عادة، في أي مدينة، من قوائم السجلات التموينية، وقد اقتُرح، بالإضافة إلى التقسيم الجغرافي إلى طبقات

ضمن المدينة، والذي تكون المعلومات حوله متوافرة عادة من حينها، اقتُرح تقسيم آخر إلى طبقات وفق الجنس والعمر. وبما أنه كان من الضروري أن تُحسب العيّنة بسرعة، وأن القوائم كانت قيد الاستخدام بصورة دائمة، فإن جدول التوزيع وفقًا للعمر والجنس لم تكن ممكنة. وعلى أي حال، فقد كان ممكنًا وبسرعة، سحب عيّنة نمطية كبيرة إلى حد معتدل وقد صُنّف كل شخص يتم سحب اسمه إلى فئة العمر ـ الجنس المناسبة. ثم اختيرت من هذا البيان الإحصائي القائمة الأصغر بكثير من الأشخاص الذين ستجري مقابلتهم.

(١٢-٢) المعاينة المضاعفة في حالة التقسيم إلى طبقات

أول من قدّم هذه النظرية هو Neyman (1938)، حيث يجري تقسيم المجتمع تقسيمًا طبقيًا إلى L من الفصول (طبقات). والعيّنة الأولى هي عيّنة عشوائية بسيطة حجمها n' . ليكن،

$W_h = N_h/N =$ نسبة عناصر المجتمع التي تقع في الطبقة h .

$w_h = n'_h/n' =$ النسبة من العيّنة الأولى التي تقع في الطبقة h .

فعندئذ تكون w_h تقديرًا لـ W_h .

والعيّنة الثانية هي عيّنة عشوائية طبقية حجمها n ، ونقيس فيها y_{hi} ، حيث تُسحب n_h من الوحدات من الطبقة h . وتكون العيّنة الثانية من الطبقة h ، عادة، عيّنة جزئية عشوائية من الـ n'_h عنصرًا الموجودة في هذه الطبقة. وهدف العيّنة الأولى هو تقدير ترجيحات الطبقات؛ أما هدف الثانية فتقدير متوسطات الطبقات $\bar{Y}_h$ ،

ومتوسط المجتمع هو $\bar{Y} = \sum W_h \bar{Y}_h$. ونستخدم كتقدير له،

$$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h \qquad (12.1)$$

والمسألة هي اختيار n' والـ n_h بحيث نجعل $V(\bar{y}_{st})$ أصغر ما يمكن، من أجل تكلفة مثبتة.

ويجب التحقق عندئذ مما إذا كان التباين الأصغري هذا أصغر مما كان يمكن بلوغه بأخذ عيّنة عشوائية بسيطة نقيس فيها y_i فقط. وعند تقديم المناقشة النظرية

نفترض أن الـ n_h هي عيّنة عشوائية جزئية من الـ n_h'. وهكذا يكون $n_h = \nu_h n_h'$، حيث $0 < \nu_h \leq 1$ ويجري اختيار الـ ν_h سلفًا. ويعني تكرار المعاينة سحب كل من العيّنتين الأولى والثانية من جديد، بحيث تكون الـ w_h والـ n_h كلها متغيرات عشوائية. والمسألة إذن هي مسألة تقسيم إلى طبقات لا نعرف فيها بالضبط حجوم الطبقات (فقرة ١٥-٢).

وتوخيًا للبساطة سنقوم بتقريبين. فنفترض أن حجم العيّنة الأولى n' كبير بكفاية بحيث يكون كل W_h موجبًا. وثانيًا عندما نأتي إلى مناقشة الاستراتيج الأمثل، فإننا نفترض أن كل قيمة مثلى نحسبها لِـ ν_h ستكون أصغر أو تساوي الواحد.

نظرية (١٢-١)

التقدير $\bar{y}_{st}$ غير منحاز.

برهان

لنأخذ أولًا المتوسط فوق العيّنات التي يكون فيها w_h مثبتًا. بما أن $\bar{y}_h$ متوسط عيّنة عشوائية بسيطة من الطبقة فإن $E(\bar{y}_h) = \bar{Y}_h$. وبالإضافة إلى ذلك عندما نأخذ المتوسط فوق مختلف الاختيارات للعيّنة الأولى فإن $E(w_h) = W_h$، باعتبار أن العيّنة الأولى نفسها هي عيّنة عشوائية بسيطة. وبالتالي،

$$E(\bar{y}_{st}) = E[E(\sum w_h \bar{y}_h | w_h)] = E(\sum w_h \bar{Y}_h) = \sum W_h \bar{Y}_h = \bar{Y} \qquad (12.2)$$

نظرية (١٢-٢)

إذا كانت العيّنة الأولى عشوائية وحجمها n'؛ والعيّنة الثانية هي عيّنة جزئية عشوائية من الأولى، حجمها $n_h = \nu_h n_h'$، حيث $0 < \nu_h \leq 1$ والـ ν_h مثبتة، فعندئذ،

$$V(\bar{y}_{st}) = S^2\left(\frac{1}{n'} - \frac{1}{N}\right) + \sum_h^L \frac{W_h S_h^2}{n'}\left(\frac{1}{\nu_h} - 1\right) \qquad (12.3)$$

حيث S^2 تباين المجتمع.

برهان

نحصل على البرهان بسهولة بوساطة الحيلة المبتكرة التالية. لنفرض أن الـ y_{hi} قد قيست في جميع وحدات العيّنة الأولى الـ n_h' الواردة من الطبقة h،

وليس فقط في العيّنة الجزئية العشوائية التي حجمها n_h . فعندئذ، وبما أن $w_h = n_h'/n'$، يكون.

$$\sum_h^L w_h \bar{y}_h' = \bar{y}'$$

متوسط عيّنة عشوائية بسيطة حجمها n' من المجتمع. وبالتالي، وبأخذ المتوسط فوق اختيارات متكررة للعيّنة التي حجمها n' ، نجد،

$$V\left(\sum_h^L w_h \bar{y}_h'\right) = S^2\left(\frac{1}{n'} - \frac{1}{N}\right) \tag{12.4}$$

إلا أن

$$\bar{y}_{st} = \sum_h^L w_h \bar{y}_h = \sum_h^L w_h \bar{y}_h' + \sum_h^L w_h (\bar{y}_h - \bar{y}_h') \tag{12.5}$$

لنفرض أن الدليل 2 يشير إلى عملية أخذ المتوسط فوق جميع العيّنات الجزئية العشوائية ذات الحجم n_h التي يمكن سحبها من n_h' وحدة معطاة. فمن الواضح أن $E_2(\bar{y}_h) = \bar{y}_h'$. والنتائج التي تتبع بصورة مباشرة هي (انظر التمرين ٢-١٦) :

$$\text{cov}\,[\bar{y}_h', (\bar{y}_h - \bar{y}_h')] = 0:$$
$$\text{cov}\,(\bar{y}_h', \bar{y}_h) = V(\bar{y}_h'): \qquad V(\bar{y}_h - \bar{y}_h') = V(\bar{y}_h) - V(\bar{y}_h') \tag{12.6}$$

وبالتالي نجد في حالة w_h مثبتة،

$$V_2\left[\sum w_h (\bar{y}_h - \bar{y}_h')\right] = \Big] = \sum w_h^2 S_h^2\left(\frac{1}{n_h} - \frac{1}{n_h'}\right) = \sum \frac{w_h S_h^2}{n'}\left(\frac{1}{\nu_h} - 1\right) \tag{12.7}$$

باعتبار أن $n_h = \nu_h n_h' = \nu_h w_h n'$.

وبأخذ المتوسط فوق توزيع الـ w_h ، الذي نحصل عليه من خلال تكرار اختيار العيّنة الأولى، نجد من (12.4) ، (12.5) و (12.7) ،

$$V(\bar{y}_{st}) = S^2\left(\frac{1}{n'} - \frac{1}{N}\right) + \sum_h^L \frac{W_h S_h^2}{n'}\left(\frac{1}{\nu_h} - 1\right) \tag{12.8}$$

نتيجة ١

يمكن التعبير عن النتيجة الأخيرة بعدد من الأشكال المختلفة. بالاستناد إلى تحليل التباين،

$$(N-1)S^2 = \sum (N_h - 1)S_h^2 + \sum N_h(\bar{Y}_h - \bar{Y})^2 \qquad (12.9)$$

وبالتالي، إذا كان $g' = (N-n')/(N-1)$ نجد لدى الضرب بـ $g'/n'N$

$$\frac{(N-n')S^2}{n'N} = S^2\left(\frac{1}{n'} - \frac{1}{N}\right) = \frac{g'}{n'}\sum (W_h - N^{-1})S_h^2 + \frac{g'}{n'}\sum W_h(\bar{Y}_h - \bar{Y})^2$$

(12.10)

وهذا يعطي استنادًا إلى (12.3)،

$$V(\bar{y}_{st}) = \sum_h^L \frac{W_h S_h^2}{n'}\left(\frac{1}{\nu_h} - 1\right) + \frac{g'}{n'}\sum_h^L (W_h - N^{-1})S_h^2 + \frac{g'}{n'}\sum_h^L W_h(\bar{Y}_h - \bar{Y})^2$$

(12.11)

وفضلاً عن ذلك، نستنتج من تعريف $g' = (N-n')/(N-1)$ أن،

$$-\frac{1}{n'} + \frac{g'}{n'} = -\frac{1}{N} + \frac{g'}{n'N} \qquad (12.12)$$

وبالتالي يمكن كتابة الحدين الثاني والثالث من $\sum W_h S_h^2$ الواردين في (12.11) واللذين معاملاهما $-1/n'$ و g'/n'، بصورة بديلة لنجد:

$$V(\bar{y}_{st}) = \sum_h W_h S_h^2\left(\frac{1}{n'\nu_h} - \frac{1}{N}\right) + \frac{g'}{n'N}\sum_h (W_h - 1)S_h^2 + \frac{g'}{n'}\sum W_h(\bar{Y}_h - \bar{Y})^2$$

(12.13)

وإذا كانت تكلفة التصنيف زهيدة فقد لا يكون من المنطقي الافتراض بأن n'/N مهمل، باعتبار أنه يمكن تصنيف نسبة غير قليلة من المجتمع. وعلى أي حال ففي معظم التطبيقات يمكن إهمال الحد الذي يحوي $g'/n'N$ في (12.13). وفي هذه الحالة تأخذ (12.13) الشكل المبسّط التالي،

$$V(\bar{y}_{st}) \doteq \sum_h^L W_h S_h^2\left(\frac{1}{n'\nu_h} - \frac{1}{N}\right) + \frac{g'}{n'}\sum_h^L W_h(\bar{Y}_h - \bar{Y})^2 \qquad (12.14)$$

وكان (Rao) قد أعطى نتائج النظرية (١٢-٢) عام 1973 .

نتيجة ٢

إذا كنا في صدد تقدير نسبة من العيّنة الثانية، نعوّض العبارتين: $S^2 = NPQ/(N-1)$

$$S_h{}^2=\frac{N_hP_hQ_h}{(N_h-1)};\qquad (\bar{Y}_h-\bar{Y})^2=(P_h-P)^2$$

في (12.3) ، (12.11) و (12.13) ، ومن أجل $n'\nu_h/N$ مهمل نجد أيضاً من (12.14) أن :

$$V(p_{st})\doteq\sum_h^L\frac{W_hP_hQ_h}{n'\nu_h}+\frac{g'}{n'}\sum_h^L W_h(P_h-P)^2 \qquad (12.15)$$

نتيجة ٣

أُعطيت في الطبعة الثانية صفحة 329 نتائج تتعلق بالحالة التي تُسحب فيها العيّنة الثانية بصورة مستقلة عن الأولى بحيث لا تعتمد الـ n_h على الـ n_h' (باستثناء ما يتعلق بالفرض $n_h\le n_h'$). وفي حالة n_h/N_h مهمل، يكون للحد الرئيس في التباين البنية نفسها كما في (12.14) ، أي

$$V(\bar{y}_{st})\doteq\sum_h^L\frac{W_h{}^2S_h{}^2}{n_h}+\frac{g'}{n'}\sum_h^L W_h(\bar{Y}_h-\bar{Y})^2 \qquad (12.16)$$

وتعمّم أبحاث قدّمها Robson (1952) ، Robson و King (1953) نظرية التقسيم إلى طبقات إلى معاينة على مرحلتين، وقد طبقاها على تقدير عدد قرّاء مجلة.

(١٢-٣) محاصّة مثلى

الهدف هو اختيار الـ n' والـ ν_h بحيث يكون $V(\bar{y}_{st})$ أصغر ما يمكن من أجل تكلفة محدّدة. لتكن c' تكلفة التصنيف لكل وحدة و c_h تكلفة قياس وحدة في الطبقة h ، فمن أجل عيّنة محدّدة، لدينا :

$$C=c'n'+\sum_h c_hn_h \qquad (12.17)$$

وبما أن الـ n_h متغيرات عشوائية، نجعل التكلفة المتوقعة لـ n' والـ ν_h التي اختيرت أصغر ما يمكن.

$$E(C)=C^*=c'n'+n'\sum c_h\nu_hW_h \qquad (12.18)$$

ومن أجل $V=V(\bar{y}_{st})$ ، تقود العلاقة (12.3) إلى،

$$n'(V+S^2/N)=(S^2-\sum W_hS_h{}^2)+\sum_h\frac{W_hS_h{}^2}{\nu_h} \qquad (12.19)$$

ولا يتضمن الجـداء $C^*(V+S^2/N)$ المقـدار n' . ويبـيّن تطبيق متراجحة كوشي ـ شوارتز على هذا الجداء أنه سيكون أصغر ما يمكن إذا كان،

$$\frac{\nu_h{}^2 c_h}{S_h{}^2} = \frac{c'}{(S^2 - \Sigma W_h S_h{}^2)} \tag{12.20}$$

من أجل كل h ، وهذا يعطي

$$\nu_h = S_h[c'/c_h(S^2 - \Sigma W_h S_h{}^2)]^{1/2} \tag{12.21}$$

ونحصل على قيمة n' من معادلة توقع الكلفة (12.18) .

وبتعـويض القيم المثـلى لِـ ν_h في العلاقـة الخـاصـة بِـ $C^*(V+S^2/N)$ نجد التبـاين الأصغري على الشكل،

$$V_{min}(\bar{y}_{st}) = \frac{1}{C^*}[\sum W_h S_h \sqrt{c_h} + (S^2 - \sum W_h S_h{}^2)^{1/2}\sqrt{c'}]^2 - \frac{S^2}{N} \tag{12.22}$$

واستخدام العلاقة (12.21) من أجل محاصّة عيّنة في التطبيق العملي، يتطلب من المعرفة بالمجتمع أكثر مما يُحتمل توافره للمعاين. ومن حسن الحظ أن من مقوّمات أخطاء التخمين أنها تعوّض بعضها البعض. وهكذا إذا كانت الـ ν_h في (12.21) مرتفعة جدًّا، فستكون n' من (12.18) منخفضة جدًّا، وسوف لا تُحدد ترجيحات الطبقات بصورة جيدة كما ينبغي لها أن تكون. إلا إننا نجد كتعويض جزئي أن متوسطات الطبقات $\bar{y}_h$ ستتحدد بدقة أكبر من دقتها تحت الحل الأمثل. وعندما يتوافر القليل من المعرفة المسبقة بترجيحات الطبقات W_h ، يقترح Srinath (1971) و Rao (1973) طريقة مختلفة قليلاً لاختيار الـ n_h، يُعتقد أنها أكثر مناعة ضد التخمينات غير الموفقة لِـ W_h .

والحالة التي تقدم أسهل مسألة محاصّة هي تلك التي يكون c_h و S_h ثابتين فيها. وعندئذ تصبح (12.21) .

$$\nu_h = \nu = \left[\left(\frac{c'}{c}\right)\frac{S_w{}^2}{(S^2 - S_w{}^2)}\right]^{1/2} = \left[\left(\frac{c'}{c}\right)\frac{1}{(\phi - 1)}\right]^{1/2} \tag{12.23}$$

حيث $\phi = S^2/S_w{}^2$ الفعالية النسبية لمحاصّة تناسبية إلى معاينة عشوائية بسيطة. وهكذا إذا خمّنا أن التقسيم إلى طبقات سيخفض $V(\bar{y})$ إلى النصف، أي أن $\phi = 2$ ، فعندئذ يكون $\nu_h = (c'/c)^{1/2}$.

مثـــال

هذا المثال لم ينشأ عن مسألة النسبة المضاعفة، إلا إنه يوضح بعض مقومات الحل. ونستخدم البيان الإحصائي لِـ Jefferson في الصفحة 168 لتقدير فدادين الذرة الصفراء فلنفرض أننا إما أن نستطيع أخذ عيّنة عشوائية بسيطة من المزارع، أو نستطيع تخصيص بعض الموارد لتصنيف المزارع إلى طبقتين وفقًا لحجم المزرعة. ولدينا معلومات مجتمعية متناسبة حول فدادين الذرة هي :

الطبقات	W_h	S_h^2	S_h	$\bar{Y}_h$
1	0.786	312	17.7	19.404
2	0.214	922	30.4	51.626
المجتمع		620		26.300

لنفرض أن $C^* = 100$، $c_h = 1 = c$ وأن S^2/N مهمل. وهذا يتضمن أنه إذا لم تُستخدم المعاينة المضاعفة فيمكننا تمويل أخذ عيّنة من $n = 100$ مزرعة مما يعطي $V(\bar{y}) = 6.20$. لتكن c' التكلفة على أساس المزرعة الواحدة لتصنيف المزارع إلى طبقة 1 (أصغر أو يساوي 160 فدانًا) وطبقة 2 (أكبر من 160 فدانًا). لنعتبر التساؤلات :

١ - من أجل أي قيم لِـ (c'/c) تجلب المعاينة المضاعفة زيادة في الدقة؟

٢ - ما هي خطة المعاينة المضاعفة المثلى إذا كان $c' = c/100$ وما هو $V(\bar{y}_{st})$ الناتج ؟

٣ - في المسألة السابقة كيف تتغير الخطة وقيمة $V(\bar{y}_{st})$ إذا خُمّنت الـ ν_h بأنها ضعف كسور المعاينة المثلى؟

والإجابات :

١ - من بيان المجتمع $\Sigma W_h S_h = 20.4$ و $(S^2 - \Sigma W_h S_h^2) = 177$ وبالتالي نجد من (12.22) :

$$V_{min}(\bar{y}_{st}) = 0.01\,(20.4 + 13.3\sqrt{c'})^2$$

وإذا أردنا لهذا أن يكون أقل من 6.20 الموافق لمعاينة عشوائية بسيطة، فعندئذ $c' < 0.11$ وبالتالي، $c'/c < 1/9$

٢ - إذا كان $c'/c = 1/100$ فعندئذ تعطي (12.22) من أجل الخطة المثلى :

$$V_{min}(\bar{y}_{st}) = 0.01(20.4 + 1.3)^2 = 4.71$$

ونلاحظ أنه إذا لم يكلف التصنيف وفقًا لحجم المزرعة شيئًا، فسنجد

$$V_{min}(\bar{y}_{st}) = 0.01(20.4)^2 = 4.16$$

وفيما يتعلق بتفاصيل الخطة مع $c'/c=1/100$ نجد من (12.21)،

$$\nu_h = S_h/133; \quad \nu_1 = 0.133; \quad \nu_2 = 0.229$$

وبما أن $\Sigma W_h \nu_h = 0.1535$ نجد، من (12.18)، أن $n'=612$. وبالتالي تكون القيم المتوقعة لـ n_2, n_1 هي 30, 64 وهكذا تكون جميع الأموال تقريبًا قد أنفقت على القياس: 6% فقط على التصنيف.

٣ ـ وإذا خمّنا $\nu_1 = 0.266, \nu_2 = 0.458$ ، فعندئذ $\Sigma W_h \nu_h = 0.307$ ومن (12.18)، $n'=315$ مما يقود إلى قيم متوقعة $n_2=31, n_1=66$ ومن (12.3) سنجد أن $V(\bar{y}_{st})$ من أجل هذه الخطة هو 4.85 أي زيادة 3% فقط فوق القيمة المثلى 4.71 في المسألة ٢ .

(١٢-٤) تقدير التباين في المعاينة المضاعفة مع التقسيم إلى طبقات

إذا كان كل من $1/n'$ و $1/N$ مهملًا بالنسبة إلى الواحد (مثلًا أصغر من 0.02)، فإن تقدير عيّنة غير منحاز تقريبًا لـ $V(\bar{y}_{st})$ المعطى في العلاقة (12.14) هو ببساطة نسخة العيّنة من هذه العلاقة،

$$v(\bar{y}_{st}) = \sum_h^L w_h s_h^2 \left(\frac{1}{n'\nu_h} - \frac{1}{N}\right) + \frac{g'}{n'} \sum_h^L w_h (\bar{y}_h - \bar{y}_{st})^2 \tag{12.24}$$

$$= \sum_h^L \frac{w_h^2 s_h^2}{n_h} - \sum_h^L \frac{w_h s_h^2}{N} + \frac{g'}{n'} \sum_h^L w_h (\bar{y}_h - \bar{y}_{st})^2 \tag{12.24$'$}$$

حيث $g'=(N-n')/(N-1)$. وستكون هذه العلاقة كافية في جميع التطبيقات تقريبًا. وعندما لا يكون $1/N$ و $1/n'$ مهملين نحتاج إلى عبارة جبرية أكثر تعقيدًا.

نظرية (١٢-٣)

تقدير العيّنة غير المنحاز لـ $V(\bar{y}_{st})$ في المعاينة المضاعفة هو،

$$v(\bar{y}_{st}) = \frac{n'(N-1)}{(n'-1)N}\left[\sum_h w_h s_h^2\left(\frac{1}{n'\nu_h}-\frac{1}{N}\right)+\frac{g'}{n'}\sum_h s_h^2\left(\frac{w_h}{N}-\frac{1}{n'\nu_h}\right)+\frac{g'}{n'}\sum_h w_h(\bar{y}_h-\bar{y}_{st})^2\right] \tag{12.25}$$

برهان

من (12.13) نجد أن الشكل العام للتباين المراد تقديره هو،

$$V(\bar{y}_{st}) = \sum_h W_h S_h^2\left(\frac{1}{n'\nu_h}-\frac{1}{N}\right)+\frac{g'}{n'N}\sum_h (W_h-1)S_h^2+\frac{g'}{n'}\sum_h W_h(\bar{Y}_h-\bar{Y})^2 \tag{12.26}$$

وبـأخذ المتوسط أولاً مع بقاء n' و w_h ثابتين، ثم فوق التغيرات في الـ w_h، نجد أن متوسط $w_h s_h^2$ في (12.25) هو $W_h S_h^2$، بينما متوسط s_h^2 is S_h^2. وسنستخدم هذه النتائج بعد المعادلة (12.31) .

وفي الحد الأخير من (12.25) ،

$$\sum w_h(\bar{y}_h-\bar{y}_{st})^2 = \sum w_h \bar{y}_h^2 - \bar{y}_{st}^2 \tag{12.27}$$

وبأخذ المتوسط أولاً من أجل w_h ثابتة،

$$E(\sum w_h \bar{y}_h^2) = \sum w_h \bar{Y}_h^2 + \sum w_h S_h^2\left(\frac{1}{\nu_h w_h n'}-\frac{1}{w_h N}\right) \tag{12.28}$$

وفضلاً عن ذلك

$$E_w E(\sum w_h \bar{y}_h^2) = \sum W_h \bar{Y}_h^2 + \sum S_h^2\left(\frac{1}{\nu_h n'}-\frac{1}{N}\right) \tag{12.29}$$

وأيضًا،

$$E(\bar{y}_{st}^2) = \bar{Y}^2 + V(\bar{y}_{st}) \tag{12.30}$$

وبطرح (12.30) من (12.29) ثم الضرب بـ g'/n' نجد،

$$\frac{g'}{n'}E\sum w_h(\bar{y}_h-\bar{y}_{st})^2 = \frac{g'}{n'}\left[\sum W_h(\bar{Y}_h-\bar{Y})^2+\sum S_h^2\left(\frac{1}{\nu_h n'}-\frac{1}{N}\right)-V(\bar{y}_{st})\right] \tag{12.31}$$

وبتعويض (12.31) لإيجاد $(n'-1)NEv(\bar{y}_{st})/n'(N-1)$ من (12.25) نجد،

$$\frac{(n'-1)N}{n'(N-1)} Ev(\bar{y}_{st}) = \left(1-\frac{g'}{n'}\right) V(\bar{y}_{st}) = \frac{(n'-1)N}{n'(N-1)} V(\bar{y}_{st})$$

وهـو المـطلوب. ونـلاحظ أن الحـدين في الـوسط في (12.25) هما من مرتبـة $1/n'N$ و $1/n'^2 v_h$. ويمكن إهمالها بالنسبة للحدود الباقية إذا كان $1/N$ و $1/n'$ مهملين. وهذا يدعم الشكل الأبسط في (12.24).

وقد أعطى Rao (1973) النتيجة (12.25) بدلالة الـ n_h و n'_h كما يلي:

$$v(\bar{y}_{st}) = \frac{N-1}{N} \sum_h \left(\frac{n_h'-1}{n'-1} - \frac{n_h-1}{N-1}\right) \frac{w_h s_h^2}{n_h} + \frac{(N-n')}{N(n'-1)} \sum_h w_h (\bar{y}_h - \bar{y}_{st})^2 \tag{12.32}$$

نتيجة

لاستخدام (12.24) في تقدير نسبة نضع p_h بدلاً من $\bar{y}_h$ و $n_h p_h q_h/(n_h-1)$ بدلاً من s_h^2.

مثـــال

في عيّنـة عشـوائية بسيطة تتضمن 374 منـزلاً من منـطقـة كبيرة، كان يشغل 292 منها عائلات بيضاء و 82 منها عائلات غير بيضاء. وعيّنة من حوالي منزل من كل أربعة منازل أعطت البيان التالي حول واقع الملكية.

	مملوكة	مستأجرة	المجموع
أبيض	31	43	74
غير أبيض	4	14	18

قدّر نسبة المنازل المستأجرة في المنطقة التي سحبت منها العيّنة والخطأ المعياري لهذا التقدير.

إذا تألفت الطبقة الأولى من المنازل التي يقطنها البيض.

$$w_1 = \frac{292}{374} = 0.78, \qquad w_2 = \frac{82}{374} = 0.22$$

$$p_1 = \frac{43}{74} = 0.58, \qquad P_2 = \frac{14}{18} = 0.78$$

$$p_{st} = w_1 p_1 + w_2 p_2 = 0.624$$
$$n' = 374, \qquad n_1 = 74, \qquad n_2 = 18$$

وقد وجدنا سابقًا أن الحد الأول فقط من (12.24) هو الحد المهم . وبالتالي ،

$$v(p_{st}) = \sum_h \frac{n_h}{(n_h - 1)} \frac{w_h p_h q_h}{\nu_h n'} = \sum_h \frac{w_h^2 p_h q_h}{(n_h - 1)} = \frac{(0.78)^2(0.2436)}{73} + \frac{(0.22)^2(0.1716)}{17}$$
$$= 0.00252$$
$$s(p_{st}) = 0.050$$

والنسبة المقدّرة للبيوت المستأجَرَة هي 0.6 ± 0.05 .

(١٢ـ٥) المعاينة
المضاعفة مع مقارنات تحليلية

تعالج الفقرة (١٥ـ١٣) تحديد حجوم العيّنة في زُمر جزئية من المجتمع وذلك من أجل مقارنات تحليلية بين L من متوسطات الزمر الجزئية ، وإذا كان الهدف جعل تباينات الفروق بين متوسطي أي زوج من الزمر الجزئية مساوية لِـV ، فقد لوحظ أن محاصّة وسطية بسيطة ، وغالبًا ما تكون محاصّة مقبولة ، هي أن نحدّد أو نجعل التباين الوسطى لِلـ$L(L\text{-}1)/2$ من الفروق

$$\frac{2}{L} \sum_i^L \frac{S_i^2}{n_i} = V \tag{12.33}$$

أصغر ما يمكن .

وفي تطبيق آخر Booth وSedransk (1969) ، شكلت الزمر الجزئية تصنيفًا عامليًا 2×2 ، وكانت المسألة تقدير تأثيرات الصفوف والأعمدة بالدقة نفسها . وفي هذه الحالة اتخذ التباين المراد تحديده أو جعله أصغر ما يمكن الشكل التالي :

$$\frac{1}{2} \sum_i^2 \sum_j^2 (\theta_{i.}^2 + \theta_{.j}^2) \frac{S_{ij}^2}{n_{ij}} = V \tag{12.34}$$

حيث الـ $\theta_{i.}$ والـ $\theta_{.j}$ ترجيحات يختارها الباحث ، والرموز (ij) تدلّ على الزمر الجزئية الأربع .

وفي المناقشة الواردة في الفقرة (١٥-١٣)، يمكن معاينة الزمر الجزئية مباشرة. وتصبح المسألة مسألة معاينة مضاعفة، عندما لا يمكن التعرّف سلفًا على عناصر الزمر الجزئية، إلا أنه يمكن بتكلفة أقل نسبيًا تصنيف الوحدات إلى زمر جزئية، وذلك من عيّنة عشوائية بسيطة تمهيدية حجمها n' . وإذا وقع n_i' من العيّنة التمهيدية في الزمرة الجزئية i ، فإننا نستخلص من الـ n_i ، الحجوم n_i الخاصة بالعيّنة النهائية التي ستمدّنا قياساتها بالمقارنات المرغوبة. وأبسط دالة تكلفة هي

$$C = c'n' + c\sum n_i = c'n' + cn \tag{12.35}$$

حيث نفترض $c' \ll c$.

ومع n_i مثبت تكون العلاقتان (12.33) و (12.34) من الشكل

$$\sum_i^L \frac{a_i^2}{n_i} = V \tag{12.36}$$

حيث نفترض أن الـ S_i^2, S_{ij}^2 وبالتالي الـ a_i^2 معروفة سلفًا.

وعندما يكون الهدف جعل V أصغر ما يمكن مع ثبات التكلفة C ، يجرّب Sedransk (1965) ، Booth و Sedransk (1969) قيمًا مختلفة لـ n' على التوالي. وفي حالة n' معطى نعرف عندئذ قيمة n من (12.35) . أما إذا كان n معطى، فتكون الـ n_i التي تجعل V أصغر ما يمكن هي $n_i = na_i/(\overline{\sum a_i})$. وعلى أي حال فقد تنشأ صعوبة عند استخدام هذه القيم لـ n_i ، وذلك بسبب كون الـ n_i' ، التي قدّمتها لنا العيّنة الابتدائية، متغيرات عشوائية. وفي بعض الزمر الجزئية يمكن أن نجد $n_i' < na_i/(\overline{\sum a_i})$ بحيث تتجاوز قيم الـ n_i ، الموافقة للقيمة الصغرى، القيمة n_i' المتوافرة لنا. وسنوضح قواعد المحاصّة التي تتناول هذه الحالة في حالة $L=3$ زمر، وذلك من Sedransk (1965). لنرقّم الزمر الجزئية وفق القيم المتصاعدة لـ $n_i'(\overline{\sum a_i})/na_i$ ولنرمز بـ $(\overline{\sum_1 a_i})$ للمجموع فوق جميع الصفوف باستثناء الأول. لنأخذ،

$$n_1 = \frac{na_1}{(\sum a_i)} \quad \text{إذا} \quad n_1' \geq \frac{na_1}{(\sum a_i)}$$

$$n_1 = n_1' \quad \text{إذا} \quad n_1' < \frac{na_1}{(\sum a_i)}$$

$$n_2 = \frac{(n-n_1)a_2}{(\sum_1 a_i)} \quad \text{إذا} \quad n_2' \geq \frac{(n-n_1)a_2}{(\sum_1 a_i)} \tag{12.37}$$

$$n_2 = n_2' \quad \text{إذا} \quad n_2' < \frac{(n-n_1)a_2}{(\sum_1 a_i)}$$

$$n_3 = n - n_1 - n_2$$

حيث $\sum_1 a_i = \sum a_i - a_1$

وهـذه القـواعد ليست كاملة، ولكنها ستغطي معظم القيم n' التي يُحتمل أن تكون قريبة من المثلى. والمبدأ هو أن نبقى أقرب ما يمكن للمحاصّة $n_i \propto a_i$. ولمزيد من التفاصيل انظر Booth و Sedransk (1969).

مثال

كمثـال ينبغي أن تؤدي المعـاينـة المضاعفة فيه أداءً حسنًا نذكر المثال التالي: $C=2000$ ، $c'=1$ ، $c=10$. والزمر الجزئية الثلاث لها حجوم نسبية $W_1 = 0.05$ ، $W_2 = 0.25$ ، $W_3 = 0.70$ ، $a_i^2 = S_i^2 = 10$ ، $(i = 1, 2, 3)$

لنعتـبر أولاً معاينة وحيدة (غير مضاعفة). وبما أن اختيار، وتصنيف، وقياس وحدة معاينة، يكلّف 11 وحدة نقدية، فيمكننا تمويل $n=182$ في معاينة غير مضاعفة.

وستتطلب المحاصّة قيمًا متساوية لِـ n_i ، ولكن قيم n_i ، في المتوسط، من معاينة غير مضاعفة فيها $n=182$ ، هي 9.1 ، 45.5 و127.4 . ومفترضين $E(1/n_i) \doteq 1/E(n_i)$ ، يكون معدّل V من معاينة غير مضاعفة مساويا تقريبًا لِـ

$$E(V) \doteq 10\left(\frac{1}{9.1} + \frac{1}{45.5} + \frac{1}{127.4}\right) = 1.40$$

وفي حالـة معـاينـة مضـاعفة تبيّن الحسابات بقيم مختلفة لِـ n' أن القيم المثلى لِـ n' قريبـة من 620 ، مما يعـطي $n=138$. ومـع $n=138$ تكـون الـ n_i المثلى 46 في كل صف. إلا إن $n'=620$ يقـدم فقط قيمـة متـوقـعـة 31 في الـصـف 1 . وبـالتـالي تعـطي قاعـدة المحـاصة (12.37) من أجل معاينة مضاعفة، في المتوسط، $n_2 = n_3 = 53.5$ و $n_1 = E(n_1') = (0.05)(620) = 31$ ، مما يقود إلى

$$E(V) \doteq 10\left(\frac{1}{31}+\frac{2}{53.5}\right) = 0.70$$

أي بتخفيض قدره 50% عن $E(V)$ في حالة معاينة غير مضاعفة .

ويتناول Rao (1973) هذه المسائل وفقًا للطريقة المستخدمة في المعاينة الطبقية. وتحـدد هذه الـطريقـة الكسر ν_i من الـ n_i' في الزمرة الجزئية i التي سيجري قياسها. وإحـدى المـيزات هي أنـــه يمـكـن تحديد الـقـيمـــة المـثـلى لِـ n' تحليليًا. ومـــع $C = c'n' + cn$ كما سبق، تكون التكلفة المتوقعة،

$$C^* = c'n' + cn' \sum \nu_i W_i \tag{12.38}$$

حيث W_i الحجم النسبي للزمـرة الجزئية i . ونفترض كما سبق أن $E(1/n_i) \doteq 1/E(n_i)$ والقيمة المتوسطة لِـ V هي،

$$E(V) = \sum_i^L \frac{a_i^2}{n' W_i \nu_i} \tag{12.39}$$

ومن متراجحة كوشي ـ شوارتز نجد أن ν_i المثلى مع بقاء n' ثابتًا، معطاة بِـ

$$n' W_i \nu_i = \frac{a_i}{(\sum a_i)} \frac{(C^* - n'c')}{c} \tag{12.40}$$

شريطة أن جميع القيم ν_i أصغـــر من أو تســاوي الـواحـد. وتعـويض $n'W_i\nu_i$ من (12.40) في (12.39) يُعطي من أجل $E(V)$ الأصغري،

$$E(V) = \frac{c(\sum a_i)^2}{C^* - n'c'} \tag{12.41}$$

وبما أن $E(V)$ في (12.41) يتناقص مع تناقص n' ، فستتناقص قيمة n' المثلى حتى تصـــل، على الأقـــل، إلى القيمـــة m_1 ، مثـلاً، التي تصبـح عنـدهـا إحـدى قيم الـ ν_1، ولنقُـل ν_i ، مسـاوية الواحد. وعند هذه النقطة لا تعود (12.40) و (12.41) قابلتين للتطبيق. ومن (12.40) تصبح ν_1 مساوية للواحد عندما،

$$n' W_1 = \frac{a_1(C^* - n'c')}{c(\sum a_i)} \tag{12.42}$$

وبحلّ (12.42) من أجل m_1، وهي قيمة n' التي يكون عندها $\nu_1=1$ ، نجد،

$$m_1 = \frac{C^*}{[c' + cW_1(\sum a_i)/a_1]} \qquad (12.43)$$

ولاختيار قيم لِـ n' أصغر من m_1 ، نضع $\nu_1 = 1$ ، ونستخدم متراجحة كوشي ـ شوارتز للحصول على الـ ν_i الباقية. ونجد من أجل $i > 1$

$$n' W_i \nu_i = \frac{a_i}{(\sum_1 a_i)} \frac{C^* - n'(c' + cW_1)}{c} \qquad (12.44)$$

حيث $\sum_1 a_i = \sum a_i - a_1$ ويكون $E(V)$ الأصغري الناتج هو،

$$E(V) = \frac{a_1^{\ 2}}{n' W_1} + \frac{c(\sum_1 a_i)^2}{C^* - n'(c' + cW_1)} \qquad (12.45)$$

ومن هذه العبارة لِـ $E(V)$ نجد أن $dE(V)/dn'$ لمشتق ينعدم من أجل،

$$n' = C^* \Big/ \left\{ (c' + cW_1) + \frac{(\sum_1 a_i)}{a_1} [cW_1(c' + cW_1)]^{1/2} \right\} \qquad (12.46)$$

والقيمــة m_2 التي يكـون عنـدهـا كل من ν_1 و ν_2 مسـاويًا للواحـد، بحيث لا تعـود (12.45) صالحة للتطبيق، هي

$$m_2 = C^* \Big/ \left[(c' + cW_1) + \frac{cW_2(\sum_1 a_i)}{a_2} \right] \qquad (12.47)$$

وهكــذا تنطبق العبـارتـان (12.45) و (12.46) فوق المـدى $m_1 \geq n' \geq m_2$ فقط. وإذا لم ينعدم $dE(V)/dn'$ من أجل $n' \geq m_2$، فإننا نحتاج إلى وضع ν_1 ، ν_2 وهلم جرًّا، مساوٍ للواحــد على التــوالي حتى نعثــر على نقــطة التحـول (نقـطة النهـاية القصـوى) لِـ $E(V)$. وعلى أي حال، ففي العديد من الحالات التي تكون فيها المعاينة المضاعفة اقتصادية، يبلغ $E(V)$ قيمته القصوى من أجل $m_1 \geq n' \geq m_2$.

مثـــال

في المثـــال الـــذي استعـــرضنـــاه في هذه الفقـرة، $C^* = 2000$، $c' = 1$، $c = 10$، و $W_i = 0.05, 0.25, 0.70$ و $a_i^2 = S_i^2 = 10$. ولدينا من (12.43) و (12.47)

$$m_1 = \frac{2000}{[1 + (10)(0.05)(3)]} = 800$$

$$m_2 = \frac{2000}{[1.5 + (10)(.25)2]} = 308$$

وفضلاً عن ذلك فإن $dE(V)/dn'$ في العلاقة (12.46) ينعدم عند،

$$n' = \frac{2000}{\{1.5 + (2)[(0.5)(1.5)]^{1/2}\}} = 619$$

وبما أن هذه القيمة تقع بين 800 و 308 فهي تعطي القيمة الصغرى المطلوبة. وتعطي العلاقة (12.44) $\nu_2 = 0.346$ و $\nu_3 = 0.124$ ومن الناحية العددية، هذا الحل هو أساسًا الحل نفسه الذي وجدناه بطريقة Sedransk ، حيث كان $n' = 620$ وقيم مماثلة لِـ n_i في الزمر الجزئية الثلاث.

ويمكن تعميم الـطريقتـين بسهولة إلى حالة تكاليف قياس تختلف باختلاف الـزمـرة الجـزئية. وهنـاك أيضًـا اقـتراحات لِـ Rao (1973) تتعلق بالحالة التي يكون فيها $E(1/n_i)$ أكبر بكثير من $1/E(n_i)$.

(١٢-٦) تقديرات انحدار

في بعض تطبيقات المعاينة المضاعفة استُخدم المتغير المساعد x_i للوصول إلى تقدير انحدار لِـ $\bar{Y}$. وفي العيّنة الأولى (كبيرة) التي حجمها n' ، نقيس فقط x_i، أما في الثـانية، وهي عيّنــة جزئية عشــوائية حجمهــا $n = \nu n' = n'/k$ حيث يجري اختيار الكسر ν سلفًا، فنقيس كلًّا من x_i و y_i . ويكون تقدير $\bar{Y}$ هو،

$$\bar{y}_{lr} = \bar{y} + b(\bar{x}' - \bar{x}) \qquad (12.48)$$

حيث $\bar{x}'$ ، $\bar{x}$ هما متوسطا الـ x_i في العيّنتين الأولى والثانية و b معامل انحدار المربعات الدنيا لِـ y_i على x_i، محسوبًا من العيّنة الثانية.

وإذا لم توضع أية فرضيات حول وجود انحدار خطي في المجتمع فسيكون $\bar{y}_{lr}$ منحازًا، كما في حالة معاينة على مرحلة واحدة (فصل ٧). ومفترضين معاينة عشوائية، وإمكانية إهمال $1/n$ و$1/n'$ بالنسبة للواحد، يمكن إعطاء تقريب لـ $V(\bar{y}_{lr})$ على الشكل:

$$V(\bar{y}_{lr}) \doteq \frac{S_y^2(1-\rho^2)}{n} + \frac{\rho^2 S_y^2}{n'} - \frac{S_y^2}{N} \qquad (12.49)$$

برهان

عند إيجاد خطأ المعاينة لـ $\bar{y}_{lr}$ في معاينة عشوائية بسيطة، برهنّا أنه إذا وضعنا بدلاً من b في $\bar{y}_{lr}$ معامل انحدار المجتمع المنتهي $B = S_{yx}/S_x^2$ ، فيكون الخطأ في التقريب من مرتبة $\frac{1}{\sqrt{n}}$ بالنسبة إلى الخطأ في $\bar{y}_{lr}$، والأمر نفسه ينطبق هنا. ولذلك فإننا ندرس تباين التقريب،

$$\tilde{y}_{lr} = \bar{y} + B(\bar{x}' - \bar{x})$$

وسيرمز الدليلان 2,1 إلى تغيرات فوق مرحلتي المعاينة الأولى والثانية، ليكن $u_i = y_i - Bx_i$ وفي المرحلة الثانية، لنعتبر العيّنة الكبيرة كمجتمع منته. فعندئذ، وباعتبار أن العيّنة الصغيرة قد سُحبت عشوائيًا من العيّنة الكبيرة، نجد،

$$E_2(\tilde{y}_{lr}) = \bar{y}' : \qquad V_2(\tilde{y}_{lr}) = \left(\frac{1}{n} - \frac{1}{n'}\right) s_u'^2$$

حيث $s_u'^2$ تباين u ضمن العيّنة الكبيرة. ونستنتج أن،

$$V(\bar{y}_{lr}) \doteq V(\tilde{y}_{lr}) = V_1(\bar{y}') + E_1\left(\frac{1}{n} - \frac{1}{n'}\right) s_u'^2 \qquad (12.50)$$

$$= \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^2(1-\rho^2) \qquad (12.51)$$

طالما أن $s_u'^2$ تقدير غير منحاز لـ $S_u^2 = S_y^2(1-\rho^2)$ وبالتالي،

$$V(\bar{y}_{lr}) \doteq \frac{S_y^2(1-\rho^2)}{n} + \frac{\rho^2 S_y^2}{n'} - \frac{S_y^2}{N} \qquad (12.49)$$

وهو المطلوب.

وكما في الفقرة (٧ـ٨)، هَبْ أننا نفترض، متّبعين Royall (1970) ، أن المجتمع المنتهي، هو نفسه، عيّنة عشوائية من مجتمع فوقي لا نهائي يصحّ فيه نموذج الانحدار الخطي. وعندئذ يصبح $\bar{y}_{lr}$ نموذج لا انحياز، ويمكن الحصول على نتائج مضبوطة لتباينه حتى من عيّنة صغيرة. ليكن نموذج الانحدار في المجتمع الفوقي،

$$y = \alpha + \beta x + \varepsilon \tag{12.52}$$

حيث تكون قيم الـ ε ، مع قيم مثبتة لـ x ، مستقلة بمتوسط يساوي الصفر وتباين $\sigma_y^2(1-\rho^2)$ ، وحيث σ_y و ρ هما الآن مَعْلَمتان في المجتمع الفوقي.

ولدى تعويض $\bar{y}$ ، $\bar{Y}$ و b من (12.52) نجد بعد عمليات جبرية بسيطة،

$$\bar{y}_{lr} - \bar{Y} = \bar{\varepsilon}_n - \bar{\varepsilon}_N + \beta(\bar{x}' - \bar{X}) + \frac{(\bar{x}' - \bar{x})\sum^{n} \varepsilon_i (x_i - \bar{x})}{\sum^{n} (x_i - \bar{x})^2} \tag{12.53}$$

وبأخذ المتوسط فوق توزيع الـ ε ، نستنتج من (12.53) أن $\bar{y}_{lr}$ نموذج لا انحياز من أجل قيم x مثبت في المجتمع المنتهي والعيّنتين. وفضلاً عن ذلك نجد من (12.53) ،

$$E[(\bar{y}_{lr} - \bar{Y})^2|x] = \sigma_y^2(1-\rho^2)\left(\frac{1}{n} - \frac{1}{N}\right) + \beta^2(\bar{x}' - \bar{X})^2 + \sigma_y^2(1-\rho^2)\frac{(\bar{x}' - \bar{x})^2}{\sum (x_i - \bar{x})^2} \tag{12.54}$$

وينشأ الحد الأخير في الطرف الأيمن من (12.54) من خطأ المعاينة لِـ b ، وهو من مرتبة $\frac{1}{n}$ بالنسبة للحدين الأولين على اليمين. وبأخذ متوسط الحدّين الأولين على اليمين فوق توزيع المقادير x الناشىء عن تكرار الاختيارات العشوائية للمجتمع المنتهي والعيّنتين، نجد،

$$EV(\bar{y}_{lr}) \doteq \sigma_y^2(1-\rho^2)\left(\frac{1}{n} - \frac{1}{N}\right) + \rho^2\sigma_y^2\left(\frac{1}{n'} - \frac{1}{N}\right) \tag{12.55}$$

$$= \frac{\sigma_y^2(1-\rho^2)}{n} + \frac{\rho^2\sigma_y^2}{n'} - \frac{\sigma_y^2}{N} \tag{12.56}$$

ولهـذه العبـارة الشكـل نفسـه كما في العبـارة (12.49) ، باستثنـاء أن $\sigma_y{}^2$ و ρ في (12.56) يشيران إلى المجتمع الفوقي .

وقد عمّم Tripathi و Khan (1967) المعاينة المضاعفة مع الانحدار إلى الحالة التي يُقاس فيها في العيّنة الثانية p من المتغيرات المساعدة x ، وتُقدّر $\bar{Y}$ بوساطة الانحدار الخطي المتعدد لـ y على هذه المتغيرات . وحيث العيّنة الثانية هي عيّنة جزئية عشوائية من الأولى، ومـع افتـراض التـوزيع الـطبيعي متعـدد المتغيرات من أجل y والمقادير x ، يعطي تعميم (12.54) إلى الحالة $p > 1$، النتيجة التالية لمتوسط التباين

$$V(\bar{y}_{lr}) = \frac{S_y{}^2(1-R^2)}{n}\left[1+\frac{(n'-n)}{n'}\,\frac{p}{(n-p-2)}\right]+\frac{R^2S_y{}^2}{n'}-\frac{S_y{}^2}{N} \qquad (12.57)$$

حيث R معامل الانحدار المتعدد بين y والمقادير x .

وقد درس Chameli Bose (1943) الحالة التي تُسحب فيها العيّنة بصورة مستقلة من العيّنة الأولى .

(١٢-٧) المحاصّة المثلى
والمقارنة مع المعاينة غير المضاعفة

من علاقة التباين (12.49) التي تفترض $\frac{1}{n}$ مهملة ، يمكن مقارنة المعاينة المضاعفة مع تقـدير انحـدار بعيّنـة عشـوائية بسيطة غير مضـاعفة ، دون تعديلات من أجل الانحدار. لدينا ،

$$V+\frac{S_y{}^2}{N}=\frac{S_y{}^2(1-\rho^2)}{n}+\frac{\rho^2S_y{}^2}{n'}: \qquad C = cn + c'n' \qquad (12.58)$$

وبالاستناد إلى متراجحة كوشي - شوارتز، يكون الجداء VC أصغر ما يمكن عندما يكون

$$\frac{cn^2}{S_y{}^2(1-\rho^2)}=\frac{c'n'^2}{\rho^2S_y{}^2} \qquad (12.59)$$

بحيث يكون $\frac{n}{n'}=\left[\frac{c'}{c}\,\frac{(1-\rho^2)}{\rho^2}\right]^{1/2}$

والتعويض في VC يعطي ،

$$(VC)_{min} = S_y{}^2(\sqrt{c(1-\rho^2)}+\sqrt{c'\rho^2})^2-\frac{CS_y{}^2}{N} \qquad (12.60)$$

وهكذا، ومع تكلفة محددة C ،

$$V_{min} = S_y^2 \frac{(\sqrt{c(1-\rho^2)} + \sqrt{c'\rho^2})^2}{C} - \frac{S_y^2}{N} \qquad (12.61)$$

وإذا كُرّست جميع الموارد بدلاً من ذلك إلى عيّنة غير مضاعفة ودون تعديلات من أجل الانحدار، فيكون حجم هذه العيّنة C/c ، ويكون تباين متوسطها،

$$V(\bar{y}) = \frac{cS_y^2}{C} - \frac{S_y^2}{N} \qquad (12.62)$$

وبالتالي يعطي الاستخدام الأمثل للمعاينة المضاعفة تبايناً أصغر، إذا كان،

$$c > (\sqrt{c(1-\rho^2)} + \sqrt{c'\rho^2})^2 \qquad (12.63)$$

ويمكن التعبير عن هذه المتراجحة بطريقتين:

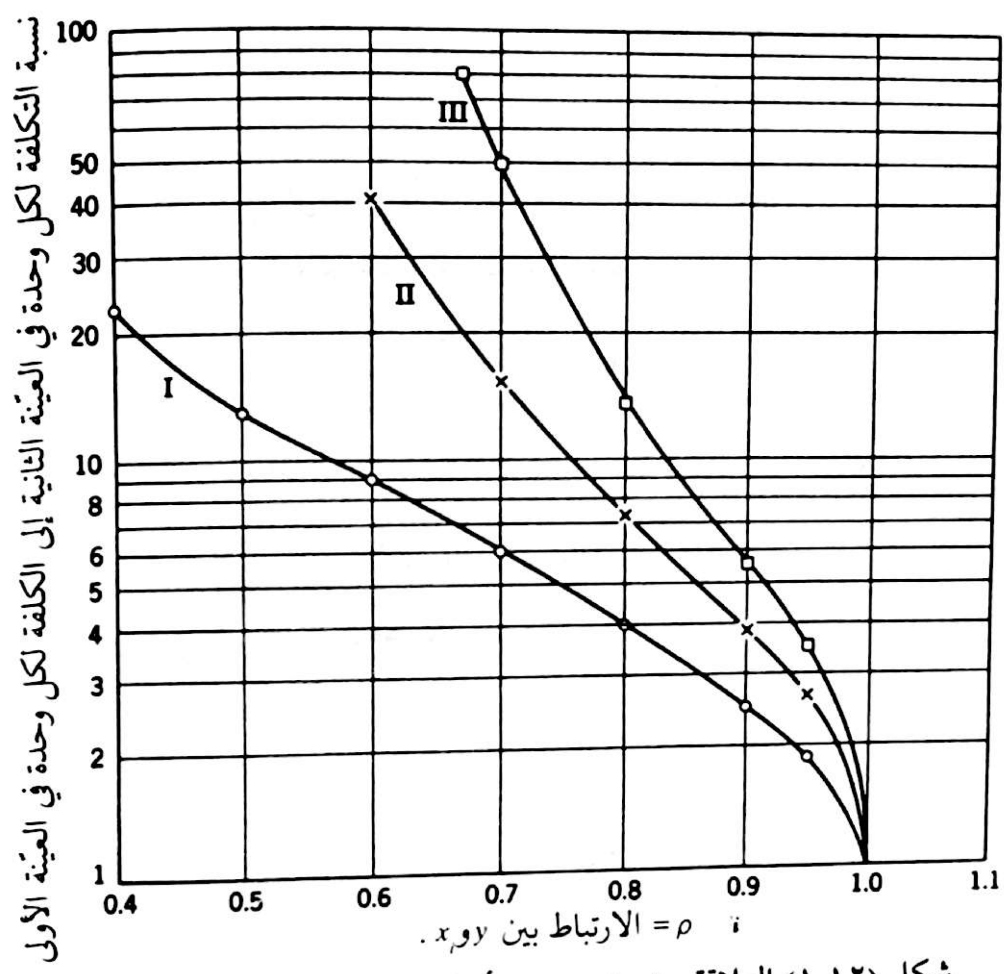
$$\frac{c}{c'} > \frac{(1+\sqrt{1-\rho^2})^2}{\rho^2} \qquad (12.64)$$

أو،

$$\rho^2 > \frac{4(c/c')}{(1+c/c')^2} \qquad (12.65)$$

وتعطي المعادلتان (12.64) و (12.65) المدَيين الحرجين لِـ c/c' في حالة ρ معطى، ولِـ ρ في حالة c/c' معطاة، مما يجعل المعاينة المضاعفة مُربحة.

ونرسم في الشكل (١٢-١) قيم النسبة c/c' (على سلّم لوغاريتمي) في مقابل ρ. والمنحنى I يمثـل العلاقة عندما تكون للمعاينتين المضاعفة وغير المضاعفة الدقة نفسها؛ والمنحني II يصحّ عندما يكون $V_{opt}=0.8V(\bar{y})$ أي عندما تعطي المعاينة المضاعفة زيادة 25% في الدقة؛ ويشير المنحني III إلى زيادة 50% في الدقة. وعلى سبيل المثال، عندما $\rho = 0.8$ ، تكون المعاينة المضاعفة في دقة المعاينة غير المضاعفة نفسها إذا كان $c/c'=4$ ، وتـعـطي 25% زيادة في الـدقـة إذا كان c/c' حوالي $7\frac{1}{2}$ ، و 50% زيادة في الدقة إذا كان c/c' حوالي 13 .



شكل (١٢-١) العلاقة بين c_u/c_n و ρ من أجل ثلاث قيم مثبتة للدقة النسبية للمعاينتين المضاعفة وغير المضاعفة .

المنحني I : المعاينتان المضاعفة وغير المضاعفة متساويتان في الدقة .

المنحني II : تعطي المعاينة المضاعفة زيادة 25% في الدقة .

المنحني III : : تعطي المعاينة المضاعفة زيادة 50% في الدقة .

وفي الاستخدامات العملية تبالغ المنحنيات في تقدير الأرباح التي نجنيها من المعاينـة المضاعفة ، لأنه إما أن نحزر أفضل القيم لِـ n و n' ، أو نقوم بتقديرها من معلومات إحصائية سابقة . وينبغي أخذ الحيطة لأخطاء إضافية في هذه التقديرات قبل أن نقرر تبنّي المعاينة المضاعفة .

ومن أجـل أي قيمـة لِـ ρ ، يوجـد حد أعلى للكسب في الدقة من المعاينة المضاعفة . ويحدث هذا عندما نحصل مجانًا على المعلومات المتعلقة بِـ $\bar{x}'(c'=0)$ والحد الأعلى للدقة النسبية هو $1/(1-\rho^2)$

(١٢ـ ٨) تقدير التباين في معاينة مضاعفة مع استخدام الانحدار

إذا أهملنا الحدود التي تحوي $1/n$،فإن $V(\bar{y}_{lr})$ معطى بالمعادلة (12.49) ،

$$V(\bar{y}_{lr}) \doteq \frac{S_y^2(1-\rho^2)}{n} + \frac{\rho^2 S_y^2}{n'} - \frac{S_y^2}{N}$$

ومع نموذج انحدار خطي تكون الكمية،

$$s_{y.x}^2 = \frac{1}{n-2}\left[\sum_{i=1}^{n}(y_i-\bar{y})^2 - b^2\sum_{i=1}^{n}(x_i-\bar{x})^2\right] \tag{12.66}$$

تقديرًا غير منحاز لِـ $S_y^2(1-\rho^2)$ وبما أن،

$$s_y^2 = \frac{\sum(y_i-\bar{y})^2}{n-1}$$

تقدير غير منحاز لِـ S_y^2 ، فينتج أن،

$$s_y^2 - s_{y.x}^2$$

تقدير غير منحاز لِـ $\rho^2 S_y^2$.

وهكذا يكون،

$$v(\bar{y}_{lr}) = \frac{s_{y.x}^2}{n} + \frac{s_y^2 - s_{y.x}^2}{n'} - \frac{s_y^2}{N} \tag{12.67}$$

تقدير عيّنة لِـ $V(\bar{y}_{lr})$.

وإذا كانت العيّنة الثانية صغيرة والحدّ في $\frac{1}{n}$ غير مهمل بالنسبة إلى الواحد، فإن التقدير المقترح للتباين في حالة عيّنات عشوائية بسيطة هو من (12.54) ،

$$v(\bar{y}_{lr}) = s_{y.x}^2\left[\frac{1}{n} + \frac{(\bar{x}'-\bar{x})^2}{\sum(x_i-\bar{x})^2}\right] + \frac{s_y^2 - s_{y.x}^2}{n'} - \frac{s_y^2}{N} \tag{12.68}$$

وهو نوع من السلالة التي تجمع بين التباين الشرطي والتباين المتوسط.

(١٢ـ٩) المقدّرات النسبة

إذا استُخدمت العيّنة الأولى للحصول على $\bar{x}$ كتقدير لِـ $\bar{X}$ في التقدير النسبة لِـ $\bar{Y}$ ، فإن مقدّر $\bar{Y}$ يكون،

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}}\bar{x}' \tag{12.69}$$

ولإيجاد التباين التقريبي نكتب،

$$\begin{aligned}\bar{y}_R - \bar{Y} &= \frac{\bar{y}}{\bar{x}}\bar{x}' - \bar{Y} \\ &= \left(\frac{\bar{y}}{\bar{x}}\bar{X} - \bar{Y}\right) + \frac{\bar{y}}{\bar{x}}(\bar{x}' - \bar{X}) \\ &= \frac{\bar{X}}{\bar{x}}(\bar{y} - R\bar{x}) + \frac{\bar{y}}{\bar{x}}(\bar{x}' - \bar{X})\end{aligned}$$

والمركبة الأولى هي خطأ التقدير النسبة العادي (فقرة ٢-١١). وعند الحصول على التباين المناسب للخطأ في الفقرة (٢-١١)، فقد وضعنا في هذا الحد واحدًا بدلاً من $\bar{X}/\bar{x}$. وللمرتبة نفسها من التقريب، نستبدل في المركبة الثانية نسبة المجتمع $R = \bar{Y}/\bar{X}$ بالعامل $\bar{y}/\bar{x}$ ، وهكذا نجد:

$$\bar{y}_R - \bar{Y} \doteq (\bar{y} - R\bar{x}) + R(\bar{x}' - \bar{X}) \tag{12.70}$$

وإذا كانت العيّنة الثانية عيّنة جزئية عشوائية من الأولى، فعندئذ،

$$E_2(\bar{y}_R - \bar{Y}) \doteq \bar{y}' - \bar{Y}: \qquad V_2(\bar{y}_R - \bar{Y}) \doteq \left(\frac{1}{n} - \frac{1}{n'}\right)s_d'^2 \tag{12.71}$$

حيث $s_d'^2$ التبـاين ضمن العيّنـة الثـانية للمتغير $d_i = y_i - Rx_i$ والآن بأخذ المتوسط فوق الاختيارات العشوائية المكررة للعيّنة الأولى نجد،

$$\begin{aligned}V(\bar{y}_R) &= V_1E_2 + E_1V_2 \\ &\doteq \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)(S_y^2 - 2RS_{yx} + R^2S_x^2)\end{aligned} \tag{12.72}$$

باعتبار أن $s_d'^2$ تقدير غير منحاز لـ $S_d^2 = S_y^2 - 2RS_{yx} + R^2S_x^2$.

وبفصل الحدّين في $\frac{1}{n}$ و$\frac{1}{n'}$ عن بعضهما، نجد،

$$V(\bar{y}_R) \doteq \frac{S_y^2 - 2RS_{yx} + R^2S_x^2}{n} + \frac{2RS_{yx} - R^2S_x^2}{n'} - \frac{S_y^2}{N} \tag{12.72'}$$

(١٢-١٠) المعاينة المتكررة من المجتمع نفسه

أصبحت عادة الاعتماد على عيّنات لجمع سلسلة مهمة من المعلومات الإحصائية التي تُنشر على فترات منتظمة، أمرًا شائعًا. ويعود هذا، جزئيًا، إلى الاعتقاد بأنه في مجتمع ديناميكي يكون التعداد الإحصائي الذي يجري في فترات غير منتظمة ذا فائدة محدودة. والمعلومات الدقيقة جدًا عن خواص مجتمع في تموز 1960 وتموز 1970 قد لا تُعين كثيرًا في تخطيط يتطلب معرفة بالمجتمع عام 1976 . وسلسلة من العيّنات الصغيرة نسبيًا في فترات زمنية سنوية أو حتى أقصر من ذلك يمكن أن تكون أكثر فائدة. وعندما نأخذ عيّنة من المجتمع نفسه (باستثناء التغيرات التي يسببها مرور الزمن) بصورة متكررة، نكون في موقع مثالي للقيام بتقديرات واقعية لكل من التكاليف والتباينات. ولتطبيق الطرق التي تقود إلى فعالية مثلى للمعاينة. وأحد الأسئلة المهمة هو الطريقة التي ينبغي بموجبها إجراء تغييرات في العيّنة مع مرور الزمن، وكم يتكرر مثل هذا التغير. وهناك العديد من الاعتبارات التي تؤثر في القرار. فقد لا يرحب الناس في تقديم معلومات من النوع نفسه مرة بعد أخرى. وقد يتأثر المستجيبون بالمعلومات التي يتلقونها في المقابلات، مما يجعلهم مع مضيّ الزمن أقل تمثيلاً للمجتمع. وعلى أي حال، يكون التعاون، أحيانًا، أفضل في المقابلة الثانية منه في المقابلة الأولى. وعندما يكون للمعلومات طابع تقني أو خصوصي (سرّي) فيمكن أن تكون المعلومات الإحصائية أكثر دقة في الزيارة الثانية منها في الزيارة الأولى.

وسنعتبر فيما تبقى من هذا الفصل مسألة استبدال العيّنة، وما يتعلق بها من مسائل القيام بتقديرات من سلسلة من العيّنات المتكررة. والموضوع ملائم لهذا الفصل لأنه يمكن الاستفادة من طرق المعاينة المضاعفة.

لتكن المعلومات الإحصائية متوافرة من سلسلة من العيّنات، فهناك ثلاثة أنواع من المقادير التي قد نرغب في تقديرها:

١ - التغير في $\bar{Y}$ من مناسبة إلى المناسبة التي تليها.

٢ - القيمة المتوسطة لِـ $\bar{Y}$ فوق جميع المناسبات.

٣ - القيمة المتوسطة لِـ $\bar{Y}$ في المناسبة الأحدث.

وفي معظم المسوح الإحصائية، يتركز الاهتمام على المتوسط الراهن (٣)،

وبصورة خاصة عندما يكون من المحتمل أن تتغير خواص المجتمع بسرعة مع الزمن، ففي حالة مجتمع يتغير ببطء مع الزمن، يمكن أن يكون المعدل السنوي (٢)، مأخوذا فوق 12 من العيّنات الشهرية، أو أربع من العيّنات الربعية، ملائمًا للاستخدامات الرئيسة. وقد تكون الحالة كذلك في دراسة لتفشّي أمراض مزمنة ومستوطنة لفترة طويلة من الزمن. ومع مرض يُظهر انتشاره تغيراً فصليًا ملحوظًا، يمكن أن تكون المعلومات الإحصائية الراهنة موضع الاهتمام الرئيس، ولكن المعدلات السنوية قد تكون مفيدة أيضًا لإجراء مقارنات بين مناطق مختلفة وأعوام مختلفة. وتقديرات التغير (١) مطلوبة بصورة رئيسة عندما نحاول دراسة تأثيرات قوى نعرف أنها فعلت فعلها في المجتمع. وعلى سبيل المثال، عند إقرار مرسوم تشريعي نفترض أنه سيشجع بناء المنازل، يكون من المفيد معرفة ما إذا كان معدل بناء البيوت الحديثة قد ازداد في العام اللاحق، (مع الاعتقاد بأن الزيادة قد لا تعود كليًا إلى صدور المرسوم التشريعي).

لنفرض أننا أحرار في تغيير تركيب العيّنة أو الاحتفاظ به، وأن الحجم الكلي للعيّنة سيبقى نفسه في جميع المناسبات، فإذا رغبنا في جعل الدقة فيمكن وضع العبارات التالية حول سياسة الاستبدال:

١ - من الأفضل، من أجل تقدير التغير، أن نحافظ على العيّنة نفسها خلال جميع المناسبات.

٢ - من أجل كل تقدير للمتوسط فوق جميع المناسبات، يُفضّل سحب عيّنة جديدة في كل مناسبة.

٣ - من أجل التقديرات الراهنة، نحصل على الدقة ذاتها سواء احتفظنا بالعيّنة نفسها، أو غيرناها في كل مناسبة. وقد يكون استبدال جزء من العيّنة في كل مناسبة أفضل بديل.

وتصح العبارتان ١ و٢ لأنه، وعلى وجه التقريب، يوجد دائمًا ارتباط إيجابي بين القياسات على الوحدة نفسها في مناسبتين متعاقبتين. وتباين تقدير التغير فوق وحدة ما هو $S_1^2+S_2^2-2\rho S_1S_2$ حيث يشير الدليلان إلى المناسبتين، وإذا قدّرنا الغير فوق وحدتين مختلفتين فإن التباين هو $S_1^2+S_2^2$. وعند تقدير المتوسط الإجمالي للمناسبتين يكون التباين $(S_1^2+S_2^2+2\rho S_1S_2)/4$ إذا احتفظنا بالوحدة نفسها، و $(S_1^2+S_2^2)/4$ إذا اخترنا وحدة جديدة.

ونتقصّى العبارة ٣ الأقل وضوحًا في فقرات لاحقة.

(١٢-١١) المعاينة في مناسبتين

لنفرض أن للعيّنة في كل من المناسبتين الحجم n نفسه، وأن التقديرات الراهنة ذات أهمية رئيسة. وقد درس Jessen (1942) سياسة الاستبدال. وللبساطة نفترض أن المعاينة العشوائية البسيطة قد استُخدمت.

ولمتوسط العيّنة الأولى تباين S_1^2/n حيث لاتوجد معلومات سابقة لنستفيد منها. وعند اختيار العيّنة الثانية، نحتفظ بـ m من وحدات العيّنة الأولى. ولا نعتبر الوحدات الـ u الباقية بل نأخذ بدلاً منها وحدات نختارها من بين الوحدات التي لم يجر اختيارها سابقًا. (m من أجل matched أي متلائم و u من أجل Un matched)

الرموز

$\bar{y}_{hu}$ = متوسط الجزء غير المتلائم في المناسبة h .

$\bar{y}_{hm}$ = متوسط الجزء المتلائم في المناسبة h .

$\bar{y}_h$ = متوسط كامل العينة في المناسبة h .

ويقدّم الجزءان المتلائم وغير المتلائم من العيّنة الثانية تقديرين مستقلين $\bar{y}_{2u}'$ ، $\bar{y}_{2m}'$ لـ $\bar{Y}_2$ كما هو مبيّن في الجدول (١٢-١). ونستخدم في الجزء المتلائم تقدير انحدار قائم على معاينة مضاعفة، حيث العيّنة الكبيرة هي العيّنة الأولى والمتغير المساعد x_i هو قيمة y_i في المناسبة الأولى. ويأتي تباين $\bar{y}_{2m}'$ من العلاقة (12.49) . لاحظ أن قيمتي m و n تقابلان قيمتي n و n' على الترتيب، المذكورتين في العلاقة (12.49) ، ونتجاهل الـ ت م م.

جدول (١٢-١) تقديرات من جزءي العيّنة المتلائم وغير المتلائم

	تقدير	تباين
غير متلائم	$\bar{y}_{2u}' = \bar{y}_{2u}$	$\dfrac{S_2^2}{u} = \dfrac{1}{W_{2u}}$
متلائم	$\bar{y}_{2m}' = \bar{y}_{2m} + b(\bar{y}_1 - \bar{y}_{1m})$	$\dfrac{S_2^2(1-\rho^2)}{m} + \rho^2\dfrac{S_2^2}{n} = \dfrac{1}{W_{2m}}$

ونجـد أفضـل تقـدير مركب لِـ $\bar{Y}_2$ بترجيح التقـديرين المستقلين وفقًا لعكس تباينيهما. وإذا كان W_{2u} وW_{2m} عكسي التباينين فهذا التقدير هو،

$$\bar{y}_2' = \phi_2 \bar{y}_{2u}' + (1-\phi_2)\bar{y}_{2m}' \qquad (12.73)$$

حيث

$$\phi_2 = \frac{W_{2u}}{W_{2u} + W_{2m}}$$

وبالاستناد إلى نظرية المربعات الدنيا، يكون تباين $\bar{y}_2'$ ،

$$V(\bar{y}_2') = \frac{1}{W_{2u} + W_{2m}}$$

ومن الجدول (١٢-١) نجد بعد التبسيط أن،

$$V(\bar{y}_2') = \frac{S_2^2(n - u\rho^2)}{n^2 - u^2\rho^2}. \qquad (12.74)$$

ونلاحظ أنه إذا كان $u=0$ (تلاؤم كامل) أو كان $u=n$ (عدم تلاؤم كامل)، فيكون لهذا التبـاين القيمـة نفسهـا S_2^2/n. ونحصـل على القيمة المثلى لِـ u ، بأخذ النهاية الصغرى لِـ (12.74) بالنسبة لتغيرات u . وهذا يعطي،

$$\frac{u}{n} = \frac{1}{1+\sqrt{1-\rho^2}}, \qquad \frac{m}{n} = \frac{\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}} \qquad (12.75)$$

وعند تعويض القيمة المثلى لِـ u في (12.74) نجد التباين الأصغري على الشكل،

$$V_{opt}(\bar{y}_2') = \frac{S_2^2}{2n}(1+\sqrt{1-\rho^2}) \qquad (12.76)$$

ويبيّن الجدول (١٢-٢)، من أجل سلسلة من قيم ρ ، النسبة المئوية المثلى التي ينبغي أن تكون متلائمة، والكسب النسبي في الدقة بالمقارنة مع عدم التلاؤم. ولا تتجاوز أفضل نسبة مئوية للتلاؤم أبدًا الـ50 بالمائة، وتتناقص باستمرار مع تزايد ρ . وعندما يكون ρ مساويًا للواحد، تقترح العلاقة $m=0$ ، وهي قيمة تقع خارج مدى فرضياتنا، طالما أننا فرضنا m كبيرًا إلى حد معقول. والإجراء الصحيح في هذه الحالة هو أن نأخذ $m=2$. والوحدتان المتلائمتان كافيتان لتحديد خط الانحدار بالضبط.

جدول (١٢-٢) النسبة المئوية المثلى للتلاؤم

ρ	النسبة المئوية المثلى للتلاؤم	النسبة المئوية للكسب في الدقة	النسبة المئوية للكسب من أجل $\frac{m}{n}=\frac{1}{3}$	النسبة المئوية للكسب من أجل $\frac{m}{n}=\frac{1}{4}$
0.5	46	7	7	6
0.6	44	11	11	9
0.7	42	17	17	15
0.8	38	25	25	23
0.9	30	39	39	39
0.95	24	52	50	52
1.0	0	100	67	75

وأفضل كسب يمكن بلوغه في الدقة هو 100 بالمائة عندما يكون $\rho = 1$. وما لم يكن ρ مرتفعًا، فالكسب يكون متواضعًا.

ومع أن النسبة المئوية المثلى للتلاؤم تتغير مع ρ، فإنه لا يمكن أن نستخدم عمليًا، من أجل جميع مفردات بيان إحصائي؛ إلا نسبة مئوية واحدة فقط. وتبين الأعمدة على يمين الجدول (١٢-٢) النسبتين المئويتين للكسب في الدقة عندما يكون ثلث أو ربع الوحدات في حالة تلاؤم. وتشكل كلتاهما تسوية جيدة، وذلك باستثناء المفردات التي يتجاوز ρ من أجلها 0.95.

ويعطي Kulldorff (1963) مناقشة مستفيضة لهذا النموذج، ويدرس الحالة التي يمكن أن تختلف فيها تكلفة قياس وحدة متلائمة في المناسبة الثانية (أي وحدة قيست سابقًا) عن تكلفة قياس وحدة غير متلائمة، ولا يفترض حجوم عيّنة متساوية في مناسبتين. وهكذا، وباستثناء ما كان من التكاليف مثبتًا، نجد أن تكلفته في المناسبة الثانية هي،

$$C_2 = mc_m + uc_u : \qquad \frac{C_2}{c_u} = m\delta + u \tag{12.77}$$

حيث $\delta = c_m/c_u$. وإذا كانت حجوم العيّنة نفسها في المناسبتين، بحيث إن $m + u = n$، فنجد النسبة المثلى لعدم التلاؤم في المناسبة الثانية بأخذ النهاية الصغرى لِـ

$$\frac{VC_2}{c_u S_2^{\,2}} = [n\delta + u(1-\delta)]\frac{(n-u\rho^2)}{(n^2-u^2\rho^2)} = [\delta + \mu(1-\delta)]\frac{(1-\mu\rho^2)}{(1-\mu^2\rho^2)} \qquad (12.78)$$

حيث $\mu = u/n$ أما V فنأخذها من (12.74) . وإذا كان $\delta<1$ ، أي أن التلاؤم أقل تكلفة، فالنسبة المثلى للتلاؤم هي عندئذ، بالطبع، أكبر من القيم المعطاة في الجدول (١٢-٢) . ويعالج أيضًا الحالة التي يُفترض فيها تساوي التكاليف في المناسبتين.

وفي بعض التطبيقات يقدّم البيان الإحصائي للمناسبة الأولى عدة متغيرات مساعدة ترتبط بـ y_2، وأحدها بالطبع سيكون y_1 كالمعتاد. وعلى سبيل المثال، عند تقدير عدد طيور الماء y_2 التي يقتلها صياد واحد في أونتاريو من 1968 إلى 1969 وجد Sen (1973 a) أن عدد طيور الماء المقتولة لكل صياد، وعدد أيام الصيد في 1967 و 1968 كان كل منهما مرتبطًا بـ y_2. وفي ذلك البحث عُمّمَ التحليل السابق إلى الحالة التي يجري فيها تعديل $\bar{y}_{2m}$ وفقًا لانحدار خطي متعدد على المتغيرات المساعدة، وحيث تكون حجوم العيّنات في المناسبتين غير متساوية. وفي حالة عيّنات كبيرة متساوية الحجم يكون التغيير الوحيد في (12.76) الخاصة بـ $V_{opt}(\bar{y}_2')$ هو أن نضع بدلاً من ρ^2 المربع R^2 لمعامل الارتباط المتعدد بين y_2 والمتغيرات المساعدة (مفترضين توزيعًا طبيعيًا بعدة متغيرات) . والنظرية المقابلة المتعلقة بالحالة التي نعدّل فيها $\bar{y}_{2m}$ وفقًا للتقدير النسبة بعدة متغيرات معطاة في Sen (1972) ، من أجل حجوم متساوية، وفي Sen (1973 b) من أجل حجوم عيّنة غير متساوية.

(١٢-١٢) المعاينة في أكثر من مناسبتين

دُرست المسألة العامة للاستبدال من قِبل Yates (1960) و Patterson (1950) وذلك بالنسبة لكل من التقديرات الراهنة وتقديرات التغيّر. وعند وجود أكثر من مناسبتين، تزداد فرص الاستخدام المرن للبيان الإحصائي. وفي المناسبة h يمكن أن تكون لدينا أجزاء من العيّنة متلائمة مع المناسبة (h-1) ، وأجزاء متناسبة مع كل من المناسبتين (h-1) و (h-2) وهلمّ جرًا. وفي محاولة تحسين التقدير الراهن، يمكن أن نجرّب انحدارًا متعددًا يتضمن جميع حالات التلاؤم مع مناسبات سابقة. ويمكن أيضًا تعديل أو تنقيح التقدير الراهن الموافق للمناسبة (h-1) ، بعد معرفة البيان الإحصائي الموافق

للمناسبة h . وفي التقدير المنقّح يمكن الاستفادة من انحدار المناسبة (h-1) على كلتا المناسبتين (h-2) و h . وذلك بفرض أنه تتوافر لنا أجزاء من العيّنة متلائمة بصورة مناسبة .

وتحوي هذه الفقرة مدخلاً إلى الموضوع . وسنقصر انتباهنا على التقديرات الراهنة التي لا نستخدم فيها الانحدار إلا على العيّنة السابقة مباشرة. ويُنتج هذا بعض الخسارة في الدقة، ولكن وباعتبار أن الارتباط ρ يتناقص عادة كلما ازدادت الفترة الزمنية بين المناسبات، فمن النادر أن تكون الخسارة في الدقة كبيرة. ونفترض هنا أن التباين S^2 ، ومعامل الارتباط ρ بين قيمتي المفردة على الوحدة نفسها في مناسبتين متتاليتين، يبقيان ثابتين في مجمل المناقشة .

وفي المناسبة الـ h لنفرض أن m_h و u_h هما عدد الوحدات المتلائمة وغير المتلائمة، على الترتيب، مع المناسبة الـ (h-1) فتقديرا $\bar{Y}_h$ التي يمكن القيام بهما معطيان في الجدول (١٢-٣)، والتغير الوحيد في الإجراءات عن المناسبة الثانية (جدول ١٢-١) هو أننا نستخدم في التعديل الانحداري للتقدير المحسوب من الجزء المتلائم، التقدير المحسّن $\bar{y}'_{h-1}$ بدلاً من متوسط العيّنة $\bar{y}_{h-1}$.

جدول (١٢-٣) تقديرات $\bar{Y}_h$ في المناسبة الـ h

	تقدير	تباين
غير متلائم	$\bar{y}_{hu}' = \bar{y}_{hu}$	$\dfrac{S^2}{u} = \dfrac{1}{W_{hu}}$
متلائم	$\bar{y}_{hm}' = \bar{y}_{hm} + b(\bar{y}'_{h-1} - \bar{y}_{h-1,m})$	$\dfrac{S^2(1-\rho^2)}{m} + \rho^2 V(\bar{y}'_{h-1}) = \dfrac{1}{W_{hm}}$

وتباين التقدير المتلائم $\bar{y}_{hm}'$ في الجدول (١٢-٣) مستخلص من المعادلة (12.49) . ونلاحظ أن (ا) m هنا تقابل n في(12.49) و(ب) نضع $\rho^2 V(\bar{y}'_{h-1})$ بدلاً من الحد $\rho^2 S_y^2/n'$ في (12.49) الذي يساوي $B^2 V(\bar{x}')$ باعتبار أن $B=\rho$ عندما يكون S ثابتًا في المناسبات المتتالية، و $\bar{y}'_{h-1}$ تقابل $\bar{x}'$ في التحليل الوارد من قبل.

وندرس الآن الدقة التي نحصل عليها إذا استخدمنا قيم m_h وu_h المثلى والترجيحات المثلى في كل مناسبة. وسنجد أن الـ m_h/n المثلى تزداد باطراد في المناسبات المتتالية، وتتقارب بسرعة من النهاية $\frac{1}{2}$.

وبالترجيح وفقًا لعكس التباين يكون التقدير الأفضل لِـ $\bar{y}_h$ هو،

$$\bar{y}_h' = \phi_h \bar{y}_{hu}' + (1-\phi_h)\bar{y}_{hm}' \qquad (12.79)$$

حيث $\phi_h = W_{hu}/(W_{hu} + W_{hm})$ وهذا يعطي،

$$V(\bar{y}_h') = \frac{1}{W_{hu} + W_{hm}} = \frac{g_h S^2}{n}$$

حيث يرمز g_h لنسبة التباين في المناسبة h إلى التباين في المناسبة الأولى. وبتعويض W_{hu}, W_{hm} من الجدول (١٢-٣)، نجد،

$$\frac{S^2}{V(\bar{y}_h')} = \frac{n}{g_h} = S^2(W_{hu} + W_{hm}) = u_h + \frac{1}{\frac{(1-\rho^2)}{m_h} + \frac{\rho^2 g_{h-1}}{n}} \qquad (12.80)$$

ونختار الآن m_h وu_h بحيث يجعلان هذه الكمية أعظم ما يمكن، وبالتالي يجعلان $V(\bar{y}_h')$ أصغر ما يمكن. وبكتابة $u_h = n - m_h$، واشتقاق الطرف الأيمن من (12.80) بالنسبة إلى m_h، نحصل على،

$$\frac{1-\rho^2}{m_h^2} = \left(\frac{1-\rho^2}{m_h} + \frac{\rho^2 g_{h-1}}{n}\right)^2$$

وهذه تعطي لدى حلّها بالنسبة لِـ $\hat{m}_h$ المثلى،

$$\frac{\hat{m}_h}{n} = \frac{\sqrt{1-\rho^2}}{g_{h-1}(1+\sqrt{1-\rho^2})} \qquad (12.81)$$

وعند تعويض هذه القيمة في (12.80) تصبح العلاقة بعد بعض العمليات الجبرية،

$$\frac{1}{g_h} = 1 + \frac{1-\sqrt{1-\rho^2}}{g_{h-1}(1+\sqrt{1-\rho^2})} \qquad (12.82)$$

ويمكن كتابة هذه العلاقة على الشكل،

$$r_h = 1 + br_{h-1}$$

حيث $r_h = 1/g_h$ و $r_1 = 1/g_1 = 1$. والاستخدام المكرر لعلاقة التتالي هذه يعطي،

$$\frac{1}{g_h} = r_h = 1 + b + b^2 + \cdots + b^{h-1} = \frac{1-b^h}{1-b}$$

حيث $b = (1-\sqrt{1-\rho^2})/(1+\sqrt{1-\rho^2})$. كما نجد من (12.82) أو بما أن $0 < b < 1$، فيكون عامل التباين الحدّي g_∞

$$g_\infty = 1 - b = \frac{2\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}} \tag{12.83}$$

وبالتالي فإن تباين $\bar{y}_h'$ ينتهي إلى،

$$V(\bar{y}_\infty') = \frac{S^2}{n}\left(\frac{2\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}}\right) \tag{12.84}$$

وأخيراً، نحصل من (12.81) على نهاية $\hat{m}_h$ على الشكل،

$$\frac{\hat{m}_\infty}{n} = \frac{\sqrt{1-\rho^2}}{g_\infty(1+\sqrt{1-\rho^2})} = \frac{1}{2}$$

وذلك بصرف النظر عن قيمة ρ .

ويبيّن الجدول (١٢-٤) النسبة المئوية المثلى للتلاؤم $100\hat{m}_h/n$، كما وجدناها في (12.81) - والتباينات الناتجة من أجل $\rho = 0.7, 0.8, 0.9, 0.95$ ومن أجل سلسلة من القيم لـ h .

جدول (١٢-٤) النسبة المئوية المثلى للتلاؤم والتباينات .

	% للتلائم $100\hat{m}_h/n$				$g_h = nV(\bar{y}_h')/S^2$			
	$\rho =$				$\rho =$			
h	0.7	0.8	0.9	0.95	0.7	0.8	0.9	0.95
2	42	38	30	24	0.857	0.800	0.718	0.656
3	49	47	42	36	0.837	0.762	0.646	0.556
4	50	49	47	43	0.834	0.753	0.622	0.515
5	50	50	49	46	0.833	0.751	0.613	0.495
∞	50	50	50	50	0.833	0.750	0.607	0.476

وعنـد الـوصـول إلى المنـاسبة الرابعة تكون النسبة المئوية المثلى للتلاؤم قريبة من 50 من أجل جميع قيم ρ المبيّنة، مع أن الجدول يشير إلى قدر أصغر من التلاؤم في المناسبتين الثانية والثالثة .وتكون التخفيضات في التباين، $(1-g_h)$، متواضعة إذا كان ρ أقل من 0.8 .

(١٢-١٣) تبسيطات وتطورات إضافية

في التـطبيق العمـلي قد يحتـاج التحليل السابق إلى تعديل. فقد افترضنا أن سياسات الاستبدال لها التكاليف وإمكانات التطبيق نفسيهما. وفي المجتمعات البشرية يُحتمـل أن تكـون التكاليف الميدانية أقل إذا احتفظنا بالوحدات نفسها في عدد من المناسبات. وإذا كنا نهتم بتقدير التغير في مجموع المجتمع أو متوسطه، فإن هذا العامل يشير أيضًا في اتجاه ملاءمة أكثر من نصف الوحدات بين مناسبة والمناسبة التي تليها.

ومن المريح أيضًا الاحتفاظ بالترجيحات ونسب التلاؤم ثابتة، بدلاً من تغييرها في كل منـاسبة. وبالتالي سنتقصّى تباين $\bar{y}_h'$ وتباين تقدير التغيّر $(\bar{y}_h'-\bar{y}_{h-1}')$ عندما نُبقي m ، u وϕ ثابتـة. ونستمـر في كتـابة $V(\bar{y}_h')=g_hS^2/n$ ، مع أن القيمة الفعلية لـ g_h ستختلف عن تلك التي رأيناها في الفقرة السابقة.

والتقدير الآن هو،

$$\bar{y}_h'=\phi\bar{y}_{hu}'+(1-\phi)\bar{y}_{hm}'$$

وبتعويض العبارات الخاصة بالتباينين (من الجدول ١٢-٣)، نجد،

$$V(\bar{y}_h')=\frac{g_hS^2}{n}=\phi^2V(\bar{y}_{hu}')+(1-\phi)^2V(\bar{y}_{hm}')$$

$$S^2\left[\frac{\phi^2}{u}+\frac{(1-\phi)^2(1-\rho^2)}{m}\right]+\frac{S^2\rho^2(1-\phi)^2g_{h-1}}{n}$$

ومنه،

$$g_h=\left[\frac{\phi^2}{\mu}+\frac{(1-\phi)^2(1-\rho^2)}{\lambda}\right]+\rho^2(1-\phi)^2g_{h-1} \qquad (12.85)$$

حيث $\mu=u/n,\ \lambda=m/n$. لنكتب هذه العلاقة على الشكل،

$$g_h = a + bg_{h-1}$$

وبالتطبيق المتكرر نجد، باعتبار $g_1=1$ ،

$$g_h = \frac{a(1-b^{h-1})}{1-b} + b^{h-1}$$

وبما أن $b = \rho^2(1-\phi)^2$ أقل من الواحد فقيمة النهاية هي ،

$$g_\infty = \frac{a}{1-b} = \frac{\lambda\phi^2 + \mu(1-\phi)^2(1-\rho^2)}{\lambda\mu[1-\rho^2(1-\phi)^2]} \qquad (12.86)$$

ويمكن إيجاد قيمة الترجيحة ϕ التي تجعل نهاية التباين أصغر ما يمكن باشتقاق (12.86) . وهذا يقود إلى معادلة تربيعية جذرها،

$$\phi_{opt} = \frac{\sqrt{1-\rho^2}[\sqrt{1-\rho^2+4\lambda\mu\rho^2} - \sqrt{1-\rho^2}]}{2\lambda\rho^2}$$

وفي الممارسة العملية، سوف لا تكون قيمة ρ معروفة بالضبط وستختلف من مفردة إلى مفردة. ويمكن عادة اختيار قيمة تشكل تسوية وسطًا. ومن الواضح أن ϕ_{opt} سوف لا تكون أقل من $\mu = u/n$، باعتبار أن الجزء المتلائم من العيّنة يعطي دقة أعلى، على أساس العنصر الواحد، من الجزء غير المتلائم. وعلى سبيل المثال، في حالة μ=0.25 ، أي أن ربع العيّنة غير متلائم، نجد أن ϕ_{opt} هي 0.216 ، 0.198 و 0.164من أجل ρ مساوٍ لـ 0.7 ، 0.8 و 0.9 ، على الترتيب. واختيار $\phi = 0.2$ قد يكون مناسبًا لهذا المدى من قيم ρ .

ولتقدير التغير، نجد،

$$V(\bar{y}_h' - \bar{y}'_{h-1}) = V(\bar{y}_h') + V(\bar{y}'_{h-1}) - 2\,\mathrm{cov}\,(\bar{y}_h'\bar{y}'_{h-1}) \qquad (12.87)$$

ولإيجاد حدّ التغاير، نلاحظ أنه إذا كانت $y_{hi}, y_{h-1,i}$ القيم الخاصة بالوحدة i في المجموعة المتلائمة في المناسبتين h و $(h\text{-}1)$ ، فإن نموذجنا هو،

$$y_{hi} = \bar{Y}_h + \rho(y_{h-1,i} - \bar{Y}_{h-1}) + e_{hi}$$

حيث الـ e_{hi} مستقلة عن الـ y . وبالتعويض نجد من هذا النموذج أن،

$$\bar{y}_{hm}' = \bar{y}_{hm} + \rho(\bar{y}'_{h-1} - \bar{y}_{h-1,m}) = \bar{Y}_h + \rho(\bar{y}'_{h-1} - \bar{Y}_{h-1}) + \bar{e}_{hm}$$

وبالتالي يكون تغاير $\bar{y}_{hm}'$ و$\bar{y}'_{h-1}$ هو $\rho V(\bar{y}'_{h-1})$. ولكن،

$$\text{cov}(\bar{y}_h'\bar{y}'_{h-1}) = \text{cov}\{[\phi\bar{y}_{hu} + (1-\phi)\bar{y}_{hm}']\bar{y}'_{h-1}\} = \rho(1-\phi)V(\bar{y}'_{h-1})$$

باعتبار أن $\bar{y}_{hu}$ مستقل عن $\bar{y}'_{h-1}$.ومن (12.87) يعطي هذا،

$$V(\bar{y}_h' - \bar{y}'_{h-1}) = \frac{S^2}{n}\{g_h + g_{h-1}[1-2\rho(1-\phi)]\} \tag{12.88}$$

ومن (12.85) و (12.87) يمكن حساب تباينات $\bar{y}_h'$ و $(\bar{y}_h' - \bar{y}'_{h-1})$ الموافقة لأية قيم لِـ m، ϕ و ρ . ويبيّن الجدول (١٢ـ٥) النسب المئوية للمكاسب الناتجة في فعالية هذه التقديرات، بالنسبة إلى التقديرات التي حصلنا عليها من عيّنات مستقلة في المناسبتين. ونسب التلاؤم $\lambda = m/n$ هي $\frac{1}{2}$ و$\frac{3}{4}$. وقد أُخذت الترجيحة ϕ مساوية لِـ 0.35 في حالة $\lambda = \frac{1}{2}$ و 0.20 في حالة $\lambda = \frac{3}{4}$.

جدول (١٢ـ٥) المكاسب في دقة فعالية التقدير الراهن $\bar{y}_h'$ وتقدير التغير $(\bar{y}_h' - \bar{y}'_{h-1})$
نسب التلاؤم : $\frac{1}{2}$ و$\frac{3}{4}$

	التقدير الراهن : النسبة المئوية للكسب = $100(1-g_h)/g_h$.							
	$\rho = 0.7$		$\rho = 0.8$		$\rho = 0.9$		$\rho = 0.95$	
h	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
2	14	10	22	14	33	19	41	22
3	16	14	30	20	52	32	67	39
4	17	15	32	24	59	40	79	52
∞	17	15	33	26	62	50	89	74
RG[a]	15	10	27	18	56	40	—	—
h	تقدير التغير: النسبة المئوية للكسب = $100(2-g'_h)/g'_h$							
2	106	153	156	233	245	399	326	565
3	113	160	170	245	277	415	365	588
4	115	160	174	251	285	424	388	603
∞	115	163	178	251	292	440	397	624
RG	101	163	166	269	381	605	—	—

a موصوفة على الصفحة ٥٠٦

وليس مما يدعو للدهشة أن تكون المعالم البارزة للجدول (١٢-٥)، هي المكاسب الكبيرة في فعالية تقديرات التغير عندما يكون ρ مساويًا على الأقل 0.7 . وفضلاً عن ذلك فإن زيادة نسبة التلاؤم من $\frac{1}{2}$ إلى $\frac{3}{4}$ تُنتج مكاسب ضخمة في فعالية تقديرات التغير، على حساب خسائر أصغر في فعالية التقديرات الراهنة. وتقترح النتائج أن الاحتفاظ بـ $\frac{2}{3}$ ، $\frac{3}{4}$ أو $\frac{4}{5}$ من مناسبة إلى الأخرى، قد يكون سياسة عملية جيدة، إذا كانت التقديرات الراهنة وتقديرات التغير جميعها مهمة.

ومقارنة المكاسب في الفعالية من أجل التقدير الراهن $\bar{y}_h'$ في الجدول (١٢-٥)، مع المكاسب المثلى في الجدول (١٢-٤)، تقترح أننا، بعد المناسبة الثانية، نخسر القليل من الدقة باستخدام ترجيحة ثابتة ونسبة تلاؤم مثبتة، وذلك ما لم يكن $\rho \geq 0.95$.

وإذا تجاوز ρ الـ 0.8 فيمكن وضع 1 بدلاً من معامل الانحدار $b = \rho$ ، مع خسارة إضافية بسيطة فقط في الدقة. وهذا يعطي تقدير $\bar{y}_h''$ من الشكل،

$$\bar{y}_h'' = \phi \bar{y}_{hu} + (1-\phi)(\bar{y}_{h-1}'' + \bar{y}_{hm} - \bar{y}_{h-1,m}) \tag{12.89}$$

وفي «المسح الجاري للمجتمع»، وهو مسح مهم يقوم به مكتب الإحصاء في الولايات المتحدة بصورة شهرية، تُستبدل ربع وحدات المرحلة الثانية كل شهر، وهكذا تبقى أسرة من العيّنة ضمن العيّنة لأربعة أشهر متتالية. وتحذف الأسرة في الأشهر الثمانية اللاحقة، إلا أنها تعاد ثانية لفترة أربعة أشهر أخرى، مما يزيد قليلاً من دقة المقارنات بين عام وآخر.

وللتقدير المركب المستخدم في هذا المسح شكل يتصل بـ (12.89) إلا أنه يختلف قليلاً:

$$\bar{y}_h'' = (1-K)\bar{y}_h + K(\bar{y}_{h-1}'' + \bar{y}_{hm} - \bar{y}_{h-1,m}) \tag{12.90}$$

حيث K عامل ترجيح ثابت. والفرق هو أن التقدير الراهن $\bar{y}_h$ ، من أجل العيّنة بمجملها، يأخذ مكان الـ $\bar{y}_{hu}$ في (12.89) . والمقادير $\bar{y}_{hm}$ ، $\bar{y}_{h-1,m}$ و $\bar{y}_h$ في (12.90) هي التقديرات النسبة، من نوع لا يخلو من التعقيد.

وتباين $\bar{y}_h''$ (ويعود إلى Bershad) معطى في Hansen ، Hurwitz

و Madow (1953) ؛ انظر أيضًا الملحق في Hansen وآخرون (1955) وتقدير التغير من شهر إلى شهر هو،

$$\bar{d}_h = \bar{y}_h'' = \bar{y}_{h-1}''$$

وبما أن الوحدات الأولية تبقى بدون تغيير فإن مركبات ما ضمن الوحدات في $V(\bar{y}_h'')$ و $V(\bar{d}_h)$ هي فقط التي تتأثر بسياسة تدوير العيّنة.

وقد درس Rao و Graham (1964) إنجاز مقدّرات مركّبة في سياسات تدويرية من هذا النوع، يبقى بموجبها مستجيب ما ضمن العيّنة لمدة r شهرًا، ثم يخرج منها لمدة m شهرًا، واستخدموا كنماذج مصور ارتباط أسّي، ومصور ارتباط خطي، يهبطان إلى الصفر. وقد درس Graham (1973) مصور ارتباط أكثر تعقيدًا ونحتاجه في حال وجود ارتباط عال بين الأشهر h و $(h\text{-}12)$.

ومكاسبهم في الفعالية في حالة $r=2,4$ و $m=\infty$ تقابل نتائج الجدول (١٢-٥) في حالة $\frac{3}{4}$، $\frac{1}{2} = \lambda$. وفي مصور ارتباط أسي يكون الارتباط بموجبه بين نتيجتين من الوحدة نفسها في المناسبتين h ، $h-i$ ، هو ρ^i تُبيّن الأسطر المقابلة لـ RG في الجدول (١٢-٥)، النسب المئوية للمكاسب في الفعالية في حالة $\rho = 0.7, 0.8, 0.9$ ومع كل ρ يستخدمون في (12.90) قيمة K المثلى في التقديرات الراهنة، وذلك عندما $h \to \infty$ وكما يبيّن الجدول (١٢-٥)، يجدون أيضًا أن المكاسب في الفعالية عند تقدير التغير أكبر بكثير في المقدّرات المركبة منها في المقدّرات الراهنة.

وفي إطار أكثر شمولاً ناقش Scott و Smith (1974) دور طرق المتسلسلات الزمنية في وضع التقديرات، في مسوح مكرّرة من أنواع مختلفة.

وفي سياسة تدوير أخرى، تُسحب عيّنة جديدة في كل مناسبة، بدون أي تلاؤم. ومع معاينات أسبوعية أو شهرية، تكون هذه الخطة مناسبة عندما ينصب الاهتمام الرئيس على التقديرات السنوية وإلى مدى أقل على التقديرات نصف السنوية أو ربع السنوية، كما في مسح يتعلق بالمرض مع التأكيد على الأمراض المزمنة، على سبيل المثال. وإذا كان الاستبيان يحصل، من أجل أي وحدة، على نتائج الشهر السابق بالإضافة إلى الشهر الجاري، فيمكن اعتبار تقديرات مركبة من الشكل،

$$\bar{y}_h' = \bar{y}_h + \phi_h(\bar{y}'_{h-1} - \bar{y}_{h-1,h}) \tag{12.91}$$

حيث $\bar{y}_h$ = تقدير مأخوذ من بيان إحصائي حديث في العيّنة الراهنة.

$\bar{y}_{h-1,h}$ = تقدير مأخوذ من البيان الإحصائي لشهر سابق في العيّنة الراهنة.

$\bar{y}'_{h-1}$ = تقدير مركب يتعلق بالشهر السابق.

وقام بالمناقشة النظرية Hansen ، Hurwitz و Madow (1953) و Woodruff (1959) ، الذين طبقوها في مسوح تتعلق بمبيعات المفرّق، و Eckler (1955) . وفي مسح لتجارة المفرّق ينطوي التقدير المركب على التقدير النسبة، فهو من الشكل،

$$\bar{y}_h'' = (1-W)\bar{y}_h + W\left(\frac{\bar{y}_h}{\bar{y}_{h-1,h}}\right)\bar{y}''_{h-1}$$

حيث W عامل ترجيح . وبما أن الارتباط من شهر إلى شهر مرتفع جدًّا، يساوي في المتوسط حوالي 0.98 ، فالمكاسب في الدقة هي مكاسب كبيرة. وبعد شهر يُحسب تقدير مركب محسّن للشهر h ، مستخدمين النتائج المتعلقة بالشهر h من العيّنة الجديدة المأخوذة في الشهر $(h+1)$.

ومن الجوهري، في هذه الطريقة، أن تكون المعلومات الإحصائية التي نحصل عليها من العيّنة الراهنة عن الشهر السابق معلومات دقيقة. وقد لا يكون الأمر كذلك عندما تعتمد المعلومات الإحصائية على الذاكرة غير المسجّلة للمستجيب، ومع ذلك فيمكن للطريقة أن تعمل بنجاح إذا كانت المعلومات من النوع الذي يسجله المستجيب بعناية بحكم العادة أو الروتين.

تمارين

١٢-١ خصص مبلغ 3000 $ لمسح إحصائي يهدف إلى تقدير نسبة. وسيكلّف المسح الرئيس 10 $ لكل وحدة معاينة جزئية. وتتوافر معلومات من أضابير بتكلفة 0.25 $ لكل وحدة معاينة جزئية، مما يمكّننا من تصنيف الوحدات إلى طبقتين من الحجم نفسه تقريبًا. إذا كانت النسبة الحقيقية 0.2 في الطبقة 1 و 0.8 في الطبقة 2 ، فقدّر القيم المثلى لِـ n ، n' ، والقيمة الناتجة $V(p_{st})$. هل تُنتج المعاينة المضاعفة كسبًا في الدقة

فوق المعاينة غير المضاعفة؟ (يمكن تجاهل النسب $n'/N, n_h/N_h$) .

(١٢-٢) من أجل W_h ، P_h في التمرين (١٢-١)، أوجد نسب التكاليف c_n/c_n' التي تكون المعاينة المضاعفة معها أكثر وفرًا من المعاينة غير المضاعفة .

(١٢-٣) يحوي مجتمع L من الطبقات ذات الحجوم المتساوية . وإذا كان V_{ran} يرمز لتباين متوسط عيّنة عشوائية بسيطة،وا V_{st} ، V_{ds} هما التباينان الموافقان لمعاينة عشوائية طبقية بمحاصّة تناسبية ، ولمعاينة مضاعفة مع تقسيم إلى طبقات، فبيّن أنه، على وجه التقريب، يكون،

$$nV_{ran} = \bar{S}_h^2 + \frac{\sum_h (\bar{Y}_h - \bar{Y})^2}{L}$$

$$nV_{st} = \bar{S}_h^2$$

$$nV_{ds} = \bar{S}_h^2 + \frac{n}{n'} \frac{\sum_h (\bar{Y}_h - \bar{Y})^2}{L}$$

حيث $\bar{S}_h^2$ هو التباين الوسطي ضمن الطبقات. (يمكن الافتراض أن كلاً من N و n' كبير بالنسبة إلى L ، كما يمكن الفرض بأن $n_h=n/L$ في المعاينة المضاعفة) .

ومنه إذا كان $(RP)_{st}$ يرمز للدقة النسبية لعينة طبقية إلى عيّنة عشوائية بسيطة، مع تعريف موافق لِـ $(RP)_{ds}$ ، فبيّن أن،

$$(\mathrm{RP})_{ds} = \frac{(\mathrm{RP})_{st}}{1 + (n/n')[(\mathrm{RP})_{st} - 1]}$$

وفي حالة $(RP)_{st}=2$ ارسم $(RP)_{ds}$ مقابل n/n' ، كم يجب أن تكون هذه النسبة صغيرة حتى تكون $(RP)_{ds}=1.9$ ؟

(١٢-٤) إذا كان $\rho = 0.8$ في المعاينة المضاعفة مع الانحدار، فكم يجب أن يكون n' بالنسبة إلى n ، إذا أردنا أن تكون الخسارة في الدقة، العائدة، إلى أخطاء المعاينة في متوسط العيّنة الكبيرة، أقل من 10 بالمائة .

(١٢-٥) في أحد تطبيقات المعاينة المضاعفة مع الانحدار، كان حجم العيّنة الصغيرة 87 وحجم العيّنة الكبيرة 300 . والحسابات التالية تتعلق بالعيّنة الصغيرة:

$$\sum (y_i - \bar{y})^2 = 17{,}283, \qquad \sum (y_i - \bar{y})(x_i - \bar{x}) = 5114, \qquad \sum (x_i - \bar{x})^2 = 3248$$

احسب الخطأ المعياري لتقدير الانحدار لِـ $\bar{Y}$.

(١٢-٦) في حالة $\rho = 0.95$ تحقَّقْ من المعلومات الإحصائية المعطاة في الجدول (١٢-٤)، والمتعلقة بما ينبغي أن تكون عليه النسبة المثلى للتلاؤم وبالكسب في الدقة بالنسبة لحالة عدم التلاؤم. احسب النسبة المئوية الموافقة للكسب في الدقة إذا احتفظنا بثلث الوحدات من المناسبة الأولى إلى الثانية، ثم احتفظنا بنصف الوحدات في كل مناسبة لاحقة.

(١٢-٧) في معاينة عشوائية بسيطة في مناسبتين، لنفرض أن التقدير في المناسبة الثانية هو وفقًا لرموز الفقرة (١٢-١١)،

$$\bar{y}_2'' = (1-\phi)(\bar{y}_1 + \bar{y}_{2m} - \bar{y}_{1m}) + \phi \bar{y}_{2u}$$

(ا) متجاهلين الـ ت م م، بيِّن أن،

$$V(\bar{y}_2'') = \frac{S^2}{n}\left\{(1-\phi)^2 \frac{[1+\mu(1-2\rho)]}{\lambda} + \frac{\phi^2}{\mu}\right\}$$

حيث $\lambda = m/n, \mu = u/n$ (ب) في حالة ρ، λ، μ معطاة أوجد قيمة ϕ التي تجعل $V(\bar{y}_2'')$ أصغر ما يمكن. بيِّن أنـه إذا تجاوز ρ النصف فإن أفضـل ترجيحـة ϕ تقع بين μ و $\mu/(1+\mu)$.

(١٢-٨) في حالة $\mu = \frac{1}{4}$، $\mu = \frac{1}{2}$، $\rho = 0.8$ و $\rho = 0.9$ ، قارن بين $V(\bar{y}_2'')$ في التمـرين السـابق وتباين تقدير الانحدار المركب الأمثل $\bar{y}_2'$ ، كما تعطيه المعادلة (12.74) .

(في $\bar{y}_2''$ خذ $\phi = 0.2$ عندما يكون $\mu = \frac{1}{4}$ و $\phi = 0.4$ عندما يكون $\mu = \frac{1}{2}$). تحقق من أجـل هذه القيم لِـ ρ أن دقـة التقـدير $\bar{y}_2''$ مسـاوية تقـريبًـا لدقة $\bar{y}_2'$ في حالة $\mu = \frac{1}{4}$ ثم في حالة $\mu = \frac{1}{2}$.

(١٢-٩) تُسحب عيِّنة مستقلة حجمها n كل شهر. ومن أجل العيِّنة المأخوذة في أي شهر، نحصل على معلومات إحصائية عن الشهر الراهن والشهر الذي يسبقه. ونضع تقديرًا مركبًا $\bar{y}_h'$ كما في (12.91) ، فقرة (١٢-١٣).

$$\bar{y}_h' = \bar{y}_h + \phi_h(\bar{y}_{h-1}' - \bar{y}_{h-1,h})$$

والنموذج هو،

$$y_{hi} = \bar{Y}_h + \rho(y_{h-1,i} - \bar{Y}_{h-1}) + e_{hi}$$

حيث e_{hi} مستقلة عن الـ y ولكل منها تباين $(1-\rho^2)$. بيّن أن (أ) :

$$\bar{y}_h' - \bar{Y}_h = \bar{e}_h + \phi_h(\bar{y}_{h-1}' - \bar{Y}_{h-1}) + (\rho - \phi_h)(\bar{y}_{h-1,h} - \bar{Y}_{h-1})$$

(ب) إذا كان $V(\bar{y}_h') = g_h S^2/n$ حيث S^2 ثابت في جميع المناسبات، فعندئذ،

$$g_h = (1-\rho^2) + \phi_h^2 g_{h-1} + (\rho - \phi_h)^2$$

(ج) القيمة المثلى $\phi_h = \rho/(1+g_{h-1})$ والقيمة المثلى الناتجة g_h هي،

$$g_h = 1 - \frac{\rho^2}{1+g_{h-1}}$$

(د) نهاية g_h هي $g_\infty = \sqrt{1-\rho^2}$. وهذه النتائج أعطاها Eckler (1955) .

(١٢-١٠) إذا كانت $E_{lr} = V(\bar{y})/V(\bar{y}_{lr})$ و $E_R = V(\bar{y})/V(\bar{y}_R)$ الفعاليات النسبية لتقدير الانحدار الخطي وللتقدير النسبة لمتوسط عيّنة عشوائية بسيطة، فبيّن أنه من أجل كل من $\bar{y}_{lr}$ و $\bar{y}_R$ تكون الفعالية النسبية لـ $\bar{y}$ في معاينة مضاعفة مع اختيار أمثل لـ n/n' (تجاهل $(1/N)$) .

$$E_{ds} = E\Big/\left(1 + \sqrt{\frac{c'}{c}}\sqrt{E-1}\right)^2$$

وبالتالي لاحظ أنه مع أي من هذه التقديرات سوف لا تكون المعاينة المضاعفة مرتفعة الفعالية ما لم تكن c/c' صغيرة (مثلاً، أصغر من 1/10) . وعلى سبيل المثال، في حالة $E=6, c/c'=1/10$ يعطي $E_{ds}=2.1$.

(١٢-١١) في معاينة في مناسبتين، لنفرض أن $S_1 = S_2 = S$ ، وأن العيّنات كبيرة بحيث إن معاملي انحدار y_{1i} على y_{2i} وانحدار y_{2i} على y_{1i} في الجزء المتلائم من العيّنات في المناسبتين يساويان ρ بتقريب جيد تمامًا. نضع تقديرًا $\bar{y}_2'$ كما في الفقرة (١٢-١٠)، وتقديرًا مماثلاً $\bar{y}_1'$ مستخدمين انحدار y_{1i} على y_{2i} ، بيّن أن،

$$V(\bar{y}_2' - \bar{y}_1') = \frac{2S^2(1-\rho)}{(n-u\rho)} \qquad \text{(i)}$$

$$V(\bar{y}_2' + \bar{y}_1') = \frac{2S^2(1+\rho)}{(n+u\rho)} \qquad \text{(ii)}$$

(إحدى طرق القيام بهذا هو التعبير عن $(\bar{y}_2' \pm \bar{y}_1')$ كدوالّ خطية في $(\bar{y}_{2m} \pm \bar{y}_{1m})$ و $(\bar{y}_{2u} \pm \bar{y}_{1u})$ وهي غير مترابطة).

وكما تقترح البداهة، لاحظ أن (i) يبلغ قيمته الصغرى عندما $u=0$ ، بينما يبلغ (ii) قيمته الصغرى عندما $u=n$.

(١٢-١٢) تحدث الحالة الأكثر ملاءمة لتطبيق الطريقة المذكورة في التمرين (١٢-١) عندما تكون النسبة الحقيقية صفرًا في الطبقة 1 . ويحدث هذا عند تقدير العدد الكلي لوحدات المجتمع التي تمتلك صفةً، تكلفة قياسها مرتفعة، وتكون هناك صفة ثانية تكلفة قياسها زهيدة، وبحيث إن الوحدات التي تمتلك الصفة الثانية هي فقط التي يمكن أن تمتلك الصفة الأولى. في عيّنة عشوائية بسيطة حجمها n' ، نُحصي عدد الوحدات m' التي تمتلك الصفة 2 . ثم نسحب عيّنة جزئية حجمها $vm' = m'/k$ من بين هذه الوحدات، ونُحصي من بينها عدد الوحدات r التي تمتلك الصفة 1 .

(ا) من النظريتين (١٢-١) و(١٢-٢)، بيّن أن $\hat{Y}_1 = Nkr/n$ تقدير غير منحاز لِـ Y_1 بتباين يساوي،

$$V(\hat{Y}_1) = \frac{N^2 P_1}{n'}\left[Q_1\left(1-\frac{n'}{N}\right) + \frac{P_2(k-1)}{P_1+P_2}\right]$$

حيث P_1 و (P_1+P_2) هي نسب المجتمع التي تمتلك الصفتين 2,1 على الترتيب. افترض أن $1/N(P_1+P_2)$ مهمل.

(ب) بفرض $c'=1, c=10$ يخمّن الباحث أن $P_1 \doteq 0.25$, $P_2 \doteq 0.15$ هل تجد المعاينة المضاعفة مربحة في هذه الحالة؟



الفصل الثالث عشر

مصادر الخطأ في المسوح الإحصائية

(١٣-١) مقدمة

تفترض النظرية المقدمة عبر الفصول السابقة بكاملها استخدام نوع من المعاينة الاحتمالية، وأن الملاحظة y_i من الوحدة i هي القيمة الصحيحة من تلك الوحدة ويبرز خطأ التقدير فقط من تغيرات المعاينة العشوائية التي توجد عند قياس n من الوحدات بدلاً من وحدات المجتمع الـ N بكاملها.

وتصحّ هذه الفرضيات بصورة جيدة تقريبًا في الأنواع الأبسط من المسوح الإحصائية التي تكون أدوات القياس فيها دقيقة، ونوعية العمل ممتازة. أما في المسوح الإحصائية المعقدة، وبصورة خاصة تلك التي تنطوي على مسائل قياس صعبة، فقد تكون هذه الفرضيات بعيدة عن الصحة. وهناك ثلاثة مصادر إضافية للخطأ يمكن تواجدها، وهي كما يلي:

١ - الفشل في قياس بعض الوحدات في العيّنة المختارة. وقد يحدث هذا نتيجة الإغفال، أو بسبب الفشل في تحديد أماكن بعض الأفراد في المجتمعات البشرية، أو رفضهم الإجابة عن الأسئلة في حال معرفة أماكنهم.

٢ - أخطاء القياس في الوحدة. فقد تكون أداة القياس منحازة أو غير دقيقة. وفي المجتمعات البشرية، قد لا يمتلك المستجيبون معلومات دقيقة أو أنهم قد يعطون أجوبة منحازة.

٣ - تندرج بعض الأخطاء عند طباعة وترميز وجدولة النتائج.

وتضطرنا مصادر الخطأ هذه إلى تعديل نظرية المعاينة المتعارف عليها. والأهداف الرئيسة لمثل هذا التعديل، هي إرشادنا إلى كيفية توزيع الموارد بين هدف تخفيض

أخطاء المعاينة العشوائية، وهدف تخفيض الأخطاء الأخرى، وأيضًا لتطوير طرق لحساب الأخطاء المعيارية وحدود الثقة تحافظ على صحتها عند وجود الأخطاء الأخرى.

(١٣-٢) تأثيرات عدم الاستجابة

سنستخدم مصطلح عدم الاستجابة للإشارة إلى الفشل في قياس بعض الوحدات في العيّنة المختارة. ومن المريح في دراسة عدم الاستجابة أن نفكر في المجتمع، وكأنه منقسم إلى «طبقتين». تحوي الأولى جميع الوحدات التي يمكن الحصول على قياساتها في حال وقوع هذه الوحدات في العيّنة، والثانية تحوي الوحدات التي لا نستطيع الحصول على قياسها. ويعتمد تركيب الطبقتين بصورة صحيحة على الطرق المستخدمة لإيجاد الوحدات، والحصول على المعلومات الإحصائية وفي مسح إحصائي تتم فيه، عند الضرورة، ثلاث زيارات على الأقل لكل بيت، ويقوم فيه مشرف، يتمتع بقدرات استثنائية على الإلحاح في المتابعة، بزيارة جميع الأشخاص الذين يرفضون إعطاء معلومات إحصائية، نقول إنه في مثل هذا المسح ستكون طبقة «عدم الاستجابة» أصغر بكثير من مسح تتم فيه محاولة واحدة فقط لكل منزل.

وتقسيم المجتمع هذا إلى طبقتين متميزتين، هو بالطبع مبالغة في التبسيط. إذ يلعب الحظ دوره في تحديد ما إذا كان سيتم إيجاد وقياس الوحدة i في عدد معطى من المحاولات. وفي تحديد أكثر كمالاً للمسألة، يمكن أن نُلحق بكل وحدة احتمالاً يمثل فرصة قياسها بطريقة ميدانية معطاة، وذلك في حال وقوعها في العيّنة.

ولا تقدم العيّنة أية معلومات حول طبقة «عدم الاستجابة» ولنرمز لها بالطبقة 2 . وقد لا يكون ذلك مهماً إذا استطعنا الافتراض بأن خواص الطبقة 2 هي خواص الطبقة 1 نفسها. وعلى أي حال، فحيثما أمكن القيام بتحقيقات تبيّن أن وحدات طبقة «عدم الاستجابة» تختلف، في الغالب، عن الوحدات التي أمكن قياسها. ويظهر في الجدول (١٣-١) توضيح لذلك. إذ نستقي المعلومات الإحصائية من معاينة تجريبية لبساتين الفاكهة في نورث كارولاينا (1946) . وقد أُرسل الاستبيان نفسه ثلاث مرات في البريد إلى أصحاب البساتين. وفيما يتعلق بأحد الأسئلة ـ عدد أشجار الفاكهة ـ كانت تتوافر معلومات كاملة عن المجتمع (1950, Finkner) .

جدول (١٣-١) الاستجابة لثلاثة طلبات في استعلام بريدي

	عدد المزارعين	النسبة المئوية من المجتمع	متوسط عدد أشجار الفاكهة لكل مزارع
الاستجابة للبريد الأول	300	10	456
الاستجابة للبريد الثاني	543	17	382
الاستجابة للبريد الثالث	434	14	340
غير المستجيبين للبُرد الثلاثة	1839	59	290
مجموع المجتمع	3116	100	329

ويتضح الانخفاض الثابت في عدد أشجار الفاكهة لكل مزارع في الاستجابات المتتالية. فهذه الأعداد هي 456 للمستجيبين للبريد الأول، 38 للبريد الثاني، 340 للبريد الثالث، و 260 لمن رفضوا الاستجابة للرسائل الثلاث. والاستجابة كانت في مجملها هزيلة. فقد قصّر ما يزيد على نصف المجتمع في إعطاء معلومات إحصائية حتى بعد ثلاث محاولات.

وسندرس الآن تأثير عدم الاستجابة على تقدير العيّنة ليكن N_1 ، N_2 عدد الوحدات في الطبقتين، وليكن $W_1 = N_1/N$ ، $W_2 = N_2/N$ أي أن W_2 هي نسبة عدم المستجيبين في المجتمع. ولنفرض أن العيّنة المسحوبة من المجتمع هي عيّنة عشوائية بسيطة. فعند إتمام العمل الميداني، تتوافر لدينا معلومات عن عيّنة عشوائية بسيطة من الطبقة 1 ، ولكن لا تتوافر لنا أية معلومات من الطبقة 2 . وبالتالي يكون الانحياز في متوسط العيّنة هو:

$$E(\bar{y}_1) - \bar{Y} = \bar{Y}_1 - \bar{Y} = \bar{Y}_1 - (W_1 \bar{Y}_1 + W_2 \bar{Y}_2)$$
$$= W_2(\bar{Y}_1 - \bar{Y}_2) \qquad (13.1)$$

ومقدار الانحياز هو جداء نسبة عدم الاستجابة في الفرق بين متوسطي الطبقتين. وبما أن العيّنة لا تقدّم أية معلومات حول $\bar{Y}_2$ ، فإن حجم الانحياز مجهول ما لم يكن ممكنًا وضع حدود لِـ $\bar{Y}_2$ من مصدر ما غير البيان الإحصائي للعيّنة. وفي الغالب، تكون الحدود التي يمكن تخصيصها. في حالة متغير مستمر، من الاتساع

بحيث تصبح عديمة الجدوى.

وبالتالي نجد، في حالة متغير مستمر، أن أية نسبة كبيرة من عدم الاستجابة تجعل وضع حدّي ثقة مفيدين لِـ $\bar{Y}$ ، بدءًا من نتائج العيّنة، أمرًا مستحيلاً عادة. ونُترك هنا في موضع الاعتماد على بعض التخمين فيما يتعلق بحجم الانحياز، وبدون أية معلومات إحصائية تدعم هذا التخمين.

وعند المعاينة من أجل تقدير نسب، تكون الحالة أسهل بقليل، باعتبار أن النسبة المجهولة P_2 في الطبقة 2 يجب أن تقع بين الصفر والواحد. وإذا كانت W_2 معروفة، فتسمح لنا حدود P_2 هذه بوضع حدود ثقة لنسبة المجتمع P . فلنفرض أننا سحبنا عيّنة عشوائية بسيطة من n وحدة، وأننا حصلنا على القياسات في n_1 من هذه الوحدات. وبفرض n_1 كبيرة بكفاية فإن 95% حدود ثقة لِـ P_1 تكون معطاة بالعلاقة،

$$p_1 \pm 2\sqrt{p_1 q_1 / n_1}$$

حيث p_1 هي نسبة العيّنة، وتجاهلنا الـ ت م م.

وعندما نحاول اشتقاق عبارة ثقة حول P ، فإننا نقف على أرض ثابتة إذا افترضنا أن $p_2 = 0$ عند إيجاد $\hat{P}_L$ وأن $p_2 = 1$ عند إيجاد $\hat{P}_U$. وهكذا يمكن أن نأخذ 95% حدود ثقة على الشكل:

$$\hat{P}_L = W_1(p_1 - 2\sqrt{p_1 q_1 / n_1}) + W_2(0) \tag{13.2}$$

$$\hat{P}_U = W_1(p_1 + 2\sqrt{p_1 q_1 / n_1}) + W_2(1) \tag{13.3}$$

ومن السهل التحقق من أن هذه الحدود محافظة، أي أن:

$$Pr(\hat{P}_L \le P \le \hat{P}_U) > 0.95$$

ويمكن تضييق هذه الحدود قليلاً من خلال مناقشة أكثر حذرًا (Cochran ، Mosteller و Tukey 1954 ، ص280) باعتبار أن P_2 لا يمكن أن تكون 0 و 1 في الوقت نفسه كما افتُرض أعلاه.

وتكون الحدود واسعة إلى حدّ مزعج إذا لم تكن W_2 صغيرة جدًا. ويبيّن الجدول (١٣-٢) القيمة المتوسطة لهذه الحدود في حالة عيّنة حجمها n=1000 وسلسلة من

قيم W_2 وp_1 . . وبما أن الحدود في المعادلتين (13.2) و (13.3) تعتمد على قيمة n_1 (عدد الاستجابات في العيّنة)، فقد أخذنا في حسابات الجدول (١٣-٢) قيمتها المتوسطة أي $n_1 = nW_1$.

وتتضح لنا الزيادة السريعة في طول فترة الثقة عندما تزداد W_2 . ومما يثير الاهتمام دراسة قيم n التي سنحتاجها للحصول على الطول نفسه لفترة الثقة كما لو أن W_2 كانت صفرًا. ويمكن القيام بذلك بسهولة عندما تكون $p_1 = 0.50$ ومن أجل $W_2 = 5$ بالمائة، يبيّن الجدول (١٣-٢) أن نصف طول فترة الثقة هو 5.6 . وحجم العيّنة المكافىء n_e تحت الفرض بأنه لا توجد أي «عدم استجابة»، معطى بالمعادلة:

$$5.6 = 2\sqrt{(50)(50)/n_e}$$

$$n_e = 320$$

جدول (١٣-٢) 95 بالمائة حدود ثقة لـ P (معبراً عنها كنسبة مئوية) وذلك عندما يكون $n=1000$

النسبة المئوية لعدم الاستجابة $100W_2$	$100p_1$ النسبة المئوية في العيّنة 5	10	20	50
0	(3.6, 6.4)	(8.1, 11.9)	(17.5, 22.5)	(46.7, 53.2)
5	(3.4, 11.1)	(7.6, 16.3)	(16.5, 26.5)	(44.4, 55.6)
10	(3.2, 15.8)	(7.2, 20.8)	(15.6, 30.4)	(42.0, 58.0)
15	(3.0, 20.5)	(6.8, 25.2)	(14.7, 34.3)	(39.6, 60.4)
20	(2.8, 25.2)	(6.3, 29.7)	(13.7, 38.3)	(37.2, 62.8)

ومن أجل W_2 تساوي 10 ، 15 و 20 بالمائة، تكون قيمة n_e هي 155 ، 90 و 60 على الترتيب. ومن الواضح أنه يجدر بنا تكريس جزء مهم من الموارد لتخفيض عدم الاستجابة.

وإذا كان معدل عدم الاستجابة في المجتمع W_2 مجهولاً، كما ستكون الحال عادة، فيمكن حساب حدود ثقة محافظة من بيان العيّنة بطريقة اقترحها أحد الطلاب. ففي حساب الحد الأدنى، نفترض أن عدم الاستجابات كافة في العيّنة كانت ستعطي إجابة سلبية. وعند حساب الحد الأعلى نفترض أن عدم الاستجابات كانت ستعطي إجابة إيجابية. وعلى سبيل المثال، لنفرض أن $n=1000$ ، $n_1=800$ و $p_1=10\%$ أي أن 80 عضوًا في العيّنة استجابوا إيجابًا، وأن معدّل عدم الاستجابة هو 20% فعندئذ، وبالنسب المئوية،

$$\hat{P}_L = 8 - 2\sqrt{(8)(92)/1000} = 6.3\%$$

$$\hat{P}_U = 28 + 2\sqrt{(28)(72)/1000} = 30.8\%$$

والحدود أعرض بقليل من تلك الخاصة بـ $p_1=10\%$ ، $W_2=20\%$ في الجدول (١٣-٢).

إذا كانت W_2 معروفة من خبرة سابقة في نوع خاص من المسوح الإحصائية، فيعطي Birnbaum و Sirken (1950b, 1950a) طريقة لإيجاد حجم العيّنة n الذي يضمن، بمخاطرة قدرها α ، خطأ مطلقًا في نسبة العيّنة لا يتجاوز مقدارًا محددًا d . ولا نفترض أية معرفة مسبقة بـ P_1 ، P_2 أو P . وإذا لم يوجد أي عدم استجابة فسنأخذ (استنادًا إلى الفقرة ٤-٤)،

$$n = t_\alpha^2 PQ/d^2 \tag{13.4}$$

حيث t_α متغير طبيعي موافق لمخاطرة أن يتجاوز الخطأ المقدار d . وبدون أية معلومات مسبقة حول P يمكن أن نأخذ $P=0.50$ كأسوأ حالة مقبولة، مما يعطي،

$$n = \frac{t_\alpha^2}{4\,d^2} \tag{13.5}$$

وبأخذ أسوأ تركيب مقبول لقيمة الانحياز $W_2(P_1\text{-}P_2)$ وقيمة النسبة p_1 ، يبيّن Birnbaum و Sirken أن قيمة n التي لا تزال تضمن خطأ أقل من d وبمخاطرة تساوي α ، هي،

$$n \doteq \frac{t_\alpha^2}{4d(d - W_2)W_1} - 1 \tag{13.6}$$

ونلاحظ أنه لا تكفي أي قيمة لـ n إذا كان $W_2>d$ وإذا كان $w_2=0$ فتُختزل هذه المعادلة إلى (13.5) باستثناء الحد 1- ، الذي يأتي من عملية تقريب في التحليل. ويبين الجدول (١٣-٣). بعض قيم n الناتجة عن طريقة Birnbaum و Sirken .

ويروي لنا الجدول القصة الحزينة نفسها التي يرويها الجدول (١٣-٢). وإذا كنا نرضى بتقدير تقريبي جدًا (d=20) فيمكن معالجة قدر من عدم الاستجابة يصل إلى 10 بالمائة بمضاعفة حجم العيّنة. وعلى أي حال، فإن أية نسبة مئوية غير قليلة لعدم

جدول (١٣-٣) أصغر قيم لـ n من أجل قيم معطاة للخطأ d ، بمخاطرة $\alpha=0.05$

النسبة المئوية لعدم الاستجابة	20	15	10	5
		d (بالنسبة المئوية)		
0	24	43	96	384
2	27	50	122	653
4	31	60	166	2000
6	36	75	255	....
8	43	99	521	....
10	53	142	...	....
15	112	...	...	....

الاستجابة تجعل من المستحيل، أو من المكلف جدًّا، أن نبلغ دقة مضمونة إلى حد كبير، من خلال زيادة حجم العيّنة بين المستجيبين.

(١٣-٣) أنواع عدم الاستجابة

نعـرض في الفقـرات القـادمـة بعض طرق معـالجة مسألة عدم الاستجابة. وكتصنيف تقريبي لأنواع عدم الاستجابة نذكر ما يلي:

١ - *عدم التغطية*. وهو الفشل في تحديد موقع بعض وحدات العيّنة أو زيارتها. وهي مشكلة تبرز في وحدات المعاينة المساحية، التي يتحتّم فيها على المعاين أن يعثر على جميع منازل جادّة معيّنة ويضع قائمة بها (وفقًا لتعريف ما). كما تنشأ أيضًا عن استخـدام قوائم غير كاملة. وأحيانًا يُسبب الطقس، أو وسائل النقل الرديئة، استحالة الوصول إلى وحدات معيّنة خلال الفترة الزمنية المخصصة للمسح.

٢ - *ليس في المنزل*. تتضمن هذه الزمرة أشخاصًا يقيمون في المنزل إلا أنهم غائبون عنه مؤقتًا. والوصول إلى الأُسر التي يعمل فيها الزوجان، أو الأُسر التي لا تتضمن أطفالاً، أصعب من الوصول إلى أُسر لديها أطفال صغار أو فيها عجّز يلتزمون البيت.

٣ - *غير قادر على الجـواب*. قد لا يمتلك المستجيب المعلومـات المطلوبة في بعض الأسئلة أو أنه لا يرحّب بإعطائها. وتشكل الصياغة الماهرة للسؤال وطريقة تقديمه نوعًا من صمام الأمان.

٤ - *«الفريق الصعب»*. الأشخاص الذين يرفضون بعناد إجراء المقابلة، أو العاجزون عن إجرائها، والغائبون عن المنزل طيلة الفترة المتوافرة للعمل الميداني، هم الذين يشكلون هذا القطاع. وهو يشكل مصدر انحياز لا يتزحزح بصرف النظر عن الجهود المبذولة لاستكمال عائداتنا من المعلومات.

إن تحرّي وقياس عدم التغطية صعب. وفي معاينة مساحية، نجد أن إحدى الطرق هي إعادة زيارة الوحدات الأولية، ووضع القوائم بعناية بحيث تخدم كوسيلة للتحقق من كمال المعلومات. وأحيانًا تقدم المقارنات بين أعداد الأشخاص أو المنازل التي أحصيناها، وبين الأعداد المتوافرة من مسح آخر، نوعًا من التنبه إلى فقدان بعضها. وعندما يكون الإطار الرئيس عبارة عن دليل يتضمن عناوين الشوارع، فيمكن دعمه بعينة مساحيّة، الهدف منها معاينة أجزاء من المدينة (بنايات جديدة) لا يغطيها الدليل بصورة مناسبة، وفي الأجزاء التي يبدو الدليل دقيقًا من أجلها، يكون الهدف هو البحث عن عناوين مفقودة من الدليل. ونجد وصفًا للمسوح التي تستخدم هذه الطرق، مع مناقشة لمسألة عدم التغطية في Kish و Hess (1958) و Woolsey (1956).

وفيما يتعلق بالغائبين عن المنزل، تكون المشكلة أسهل في مسوح يستطيع فيها أي بالغ الإجابة عن الأسئلة، مما هي في مسوح نريد فيها مقابلة شخص بالغ واحد نختاره عشوائيًا. ونفضل عادة بالغًا بمفرده، إذا كان المسح من النوع الذي يتضمن أفرادًا، لا يستطيع أحدهم أن يجيب بدقة نيابة عن آخر؛ أو إذا توقعنا ارتباطات عالية ما بين المعلومات ضمن المنزل الواحد، بحيث يكون قياس أكثر من شخص واحد في كل أسرة غير اقتصادي. وفي هذا المجال، طوّر Kish (1949) طريقة مفيدة لاختيار شخص واحد من أسرة. وقد أُعدّت الخطة الأصلية لأُسر تتضمن في حدود ستة أشخاص مؤهلين للمقابلة، إلا أنه يمكن تكييف الأسلوب بحيث يناسب أُسرًا أصغر أو أكبر.

ويضع المعاين على الجدول قائمة بالأشخاص المؤهلين في الأسرة ثم يرقّمهم: الذكور أولاً مرتبين من الأعمر إلى الأقل عمرًا، ثم الإناث من الأعمر إلى الأقل عمرًا. وتُطبع على كل جدول واحدة من مجموعات التعليمات المذكورة في الجدول (١٣-٤).

ولكل شخص مؤهل في أسرة من حجم معطى الفرصة نفسها في أن يقع عليه الاختيار، باستثناء مبالغة طفيفة في تمثيل البالغَين الثالث والخامس من أُسر حجمها 5 . وبما أن المستجيبين الذكور يتركّزون في الجداول A ، B و C فيمكن للمعاين أن يخصص زيارات مسائية إلى أسر من هذا النوع .

جدول (١٣-٤) تعليمات لاختيار مستجيب واحد

إذا كان عدد البالغين في الأسر هو						رقم الجدول	التكرار النسبي للاستخدام
6	5	4	3	2	1		
اختر البالغ رقم							
1	1	1	1	1	1	A	1/6
2	2	1	1	1	1	B1	1/12
2	2	2	1	1	1	B2	1/12
3	3	2	2	1	1	C	1/6
4	4	3	2	2	1	D	1/6
5	3	3	3	2	1	E1	1/12
5	5	4	3	2	1	E2	1/12
6	5	4	3	2	1	F	1/6

(١٣-٤) الزيارات المتكررة

إحدى التقانات المتعارف عليها هي تحديد عدد الزيارات المتكررة، أو العدد الأصغر منها. الذي يجب القيام به لأي وحدة قبل تركها على أساس أنه «لا يمكن الوصول إليها». ويعطي Stephan و Mc Carthy (1958) بيانًا إحصائيًا من عدد من المسوح حول النسبة المئوية من مُجمل العيّنة التي تمّ الحصول عليها في كل زيارة. ومتوسط النتائج معروض في الجدول (١٣-٥).

وفي المسوح التي يمكن فيها لأي بالغ في المنزل أن يجيب عن الأسئلة حصلت الزيارة الأولى على حوالي 70% من العيّنة، والزيارتان الأولى والثانية على 87% . وتتضح زيادة تكلفة المعاينة عند مقابلة شخص يجري اختياره عشوائيًا، وتُنتج الزيارة الأولى 37% فقط من المقابلات المطلوبة. ويعكس النجاح الملحوظ للزيارة الثانية عمل المعاين في تقصيه سلفًا عن الوقت الذي يكون فيه المستجيب المرغوب في المنزل ومتوافرًا.

جدول (١٣-٥) عدد الزيارات المطلوبة للوصول إلى مقابلات مكتملة
النسبة المئوية للعيّنة التي جرى الاتصال بها

المستجيب	الزيارة الأولى	الزيارة الثانية	الزيارة الثالثة وما بعدها	النسبة المئوية لعدم الاستجابة	المجموع
أيّ بالغ*	70	17	8	5	100
بالغ عشوائي	37	32	23	8	100

وقد نُشر القليل عن التكاليف النسبية للزيارات التي تلي الزيارة الأولى. ومن المتوقع أن تكون الزيارات التالية أكثر تكلفة للمقابلة الكاملة، باعتبار أن المنازل أكثر تبعثرًا من حيث مواقعها في المنطقة المخصصة للمعاين، وأن سكان هذه المنازل هم أناس يُحتمل أن يكون الوقت الذي يقضونه خارج المنزل فوق المتوسط. ومن الخبرة البريطانية يقترح Durbin (1954) أن الزيارات التالية قد تكون أقل تكلفة مما هو منتظر. وتبيّن الأرقام التالية (جدول ١٣-٦) تقدير التكاليف النسبية لكل مقابلة كاملة (أي المال المصروف على الزيارات الـ i مقسومًا على عدد المقابلات الجديدة التي حصلنا عليها) لكل من الزيارات الأولى حتى الخامسة، وذلك من دراسة خاصة لـ Durbin و Stuart (1954).

جدول (١٣-٦) التكلفة النسبية لكل مقابلة كاملة جديدة عند الزيارة

الزيارة	1	2	3	4	5
التكلفة النسبية	100	112	127	151	250

ويتطلب تقدير هذه التكاليف الحذر. وإذا لم يكن المستجيب المرغوب في المنزل عند الزيارة الأولى، فقد يقضي المعاين بعض الوقت، وهو يستفسر عن الوقت الذي يكون فيه هذا الشخص في المنزل، ويحاول تحديد موعد مبدئي. وعند حساب التكلفة ينبغي تخصيص مثل هذا الوقت لحساب الزيارة الثانية بدلاً من الزيارة الأولى غير الناجحة.

• تحت عنوان أي بالغ وضعنا مسَحين كان المستجيب فيهما زوجة لا تعمل وميكانيكي مزرعة، على الترتيب.

والقياس الأكثر فائدة هو متوسط التكلفة لكل مقابلة كاملة محسوبًا فوق جميع المقابلات التي جرت منذ الزيارة الأولى وحتى الزيارة i . وتعطي هذه الأرقام التكاليف النسبية للحصول على n من المقابلات الكاملة، وذلك عندما نصرّ على i من الزيارات قبل أن يَستسلم المعاين. ولكي نحسب هذه الأرقام يجب أن نعرف عدد المقابلات التي تمّت في كل زيارة. وفي الجـدول (١٣ـ٧) جرت هذه الحسابات تحت مجموعتين من الفروض. الأولى تمثل مسوحًا يمكن فيها لأي بالغ أن يجيب عن الأسئلة، والثانية تلك التي تتطلب بالغًا نختاره عشوائيًا. وقد أُخذ البيان المتعلق بعدد المقابلات التي تمّت من الجدول (١٣ـ٥).

وتفاصيل الحساب مبيّنة في حالة الفرض الأول فقط، وطريقة الحساب تبقى نفسها بالضبط في حالة الفرض الثاني. ويشير الرمز n_o إلى الحجم الأصلي للعيّنة.

والإصرار على الزيارات حتى الثالثة يكلّف على أساس المقابلة الكاملة 4% فقط زيادة على تكلفة زيارة بمفردها، إذا كان أي بالغ مستجيبًا مقبولاً، و 10% فقط زيادة إذا كان من الضروري مقابلة بالغ نختاره عشوائيًا. والحد الذي يمكن أن نضفي فيه على هذه النتائج صفة النموذجية أمر غير معروف، إلا أن الطريقة تقدم تقديرات واقعية لتكلفة الإصرار على تكرار الزيارة، إذا تجمعت لدينا المعلومات الضرورية عن التكلفة وحجم العيّنة. وهناك أيضًا عامل الزمن فتكرار الزيارات يؤخر النتائج النهائية.

جدول (١٣ـ٧) التكاليف النسبية لكل مقابلة كاملة حتى الزيارة i

Re المستجيب = أي بالغ ult I

بالغ «عشوائي»		حتى الزيارة i			عند الزيارة i			
التكلفة لكل مقابلة	عدد المقابلات	التكلفة لكل مقابلة	التكلفة الاجمالية	العدد الكلي للمقابلات	تكلفة المقابلات	عدد المقابلات	التكلفة النسبية	الزيارة
100	$0.37n_o$	100	$70n_o$	$0.70n_o$	$70n_o$	$0.70n_o$	100	1
106	$0.32n_o$	102	$89.04n_o$	$0.87n_o$	$19.04n_o$	$0.17n_o$	112	2
110	$0.16n_o$	104	$97.93n_o$	$0.94n_o$	$8.89n_o$	$0.07n_o$	127	3
114	$0.09n_o$	106	$103.97n_o$	$0.98n_o$	$6.04n_o$	$0.04n_o$	151	4
122	$0.06n_o$	109	$108.97n_o$	$1.00n_o$	$5.00n_o$	$0.02n_o$	250	5

(١٣-٥) نموذج رياضي لتأثيرات تكرار الزيارة

طوّر Deming (1953) نموذجًا رياضيًا مرنًا ومفيدًا كي يدرس بمزيد من التفصيل نتائج مختلفة في مسألة تكرار الزيارة. وقد قُسم المجتمع إلى r صفًا وفقًا لاحتمال أن المستجيب سيكون في المنزل. ليكن

w_{ij} = احتمال الوصول إلى المستجيب في الصف j في الزيارة i أو قبلها.

P_j = نسبة عناصر المجتمع الواقعة في الصف j.

μ_j = متوسط المفردة في الصف j.

σ_j^2 = تباين المفردة في الصف j.

وللبساطة نفترض $w_{ij}>0$ في جميع الصفوف، مع أنه يمكن بسهولة تكييف الطريقة لتتضمن أشخاصًا يستحيل الوصول إليهم. وإذا كان $\bar{y}_{ij}$ المتوسط للأشخاص في الصف j الذين قوبلوا في الزيارة i أو قبلها، فنفترض أيضًا أن $E(\bar{y}_{ij}) = \mu_j$.

ومتوسط المجتمع الصحيح للمفردة هو،

$$\bar{\mu} = \sum_j p_j \mu_j \qquad (13.7)$$

لنعتبر تركيب العيّنة بعد i من الزيارات. فيمكن تصنيف الأشخاص في العيّنة إلى $(r+1)$ صفًا كما يلي: في الصف الأول وقوبلوا؛ في الصف الثاني وقوبلوا؛ وهلم جرًا. ويتضمن الصف $(r+1)$ جميع أولئك الذين لم تجر مقابلتهم بعد i من الزيارات. وإذا تجاهلنا الـ ت م م، فإن الأعداد الواقعة في هذه الصفوف الـ $(r+1)$ تتوزع وفقًا للتوزيع متعدد الحدود،

$$[w_{i1}p_1 + w_{i2}p_2 + \cdots + w_{ir}p_r + (1 - \sum w_{ij}p_j)]^{n_0}$$

حيث n_0 الحجم الابتدائي للعيّنة.

ونستنتج أن n_i عدد الذين قوبلوا خلال i من الزيارات يتوزع وفق التوزيع الثنائي، حيث عدد التكرارات n_0 واحتمال النجاح $\sum w_{ij}p_j$. وبالتالي،

$$E(n_i) = \text{العدد المتوقع للمقابلات بعد } i \text{ زيارة} = n_0 \sum_j^r w_{ij} p_j \qquad (13.8)$$

ومع بقاء n_i ثابتًا، يتبع عدد المقابلات التي حصلنا عليها $(j=1,2\ldots,r)$ التوزيع

متعدد الحدود باحتمال $w_{ij}p_j/\sum w_{ij}p_j$ ونستنتج من ذلك،

$$E(n_{ij}|n_i) = \frac{n_i w_{ij} p_j}{\sum w_{ij} p_j}$$

وبالتالي، إذا كان $\bar{y}_i$ متوسط العيّنة التي حصلنا عليها بعد i من الزيارات، فعندئذ،

$$E(\bar{y}_i|n_i) = E\left(\frac{\sum n_{ij}\bar{y}_{ij}}{n_i}\right) = \frac{\sum n_i w_{ij} p_j \mu_j}{n_i \sum w_{ij} p_j} = \frac{\sum w_{ij} p_j \mu_j}{\sum w_{ij} p_j} = \bar{\mu}_i \qquad (13.9)$$

وبما أن هذه النتائج لا تعتمد على n_i فالمتوسط غير الشرطي لـ $\bar{y}_i$ هو أيضًا $\bar{\mu}_i$ والانحياز في تقدير $\bar{y}$ هو إذن $(\bar{\mu}_i - \bar{\mu})$.

ويمكن، بصورة مماثلة، إيجاد التباين الشرطي لـ $\bar{y}_i$ من أجل n_i معطى:

$$V(\bar{y}_i|n_i) = \frac{\sum_j' w_{ij} p_j [\sigma_j^2 + (\mu_j - \bar{\mu}_i)^2]}{n_i \sum_j' w_{ij} p_j} \qquad (13.10)$$

ومتجاهلين الحدود من مرتبة $\frac{1}{n_i^2}$ يمكن أن نجد، بصورة تقريبية، التباين غير الشرطي وذلك بأن نضع بدلاً من n_i في (13.10) قيمتها المتوقعة مأخوذة من (13.8) .

وأخيراً فإن متوسط مربعات الخطأ للتقدير الذي نحصل عليه بعد i زيارة هو،

$$\text{MSE}(\bar{y}_i|i) = V(\bar{y}_i|i) + (\bar{\mu}_i - \bar{\mu})^2 \qquad (13.11)$$

ويجب أيضًا أن نأخذ تكلفة القيام بـ i زيارة في الاعتبار. والعدد المتوقع للمقابلات الجديدة التي نحصل عليها في الزيارة k هو $\sum (w_{kj} - w_{k-1,j})p_j$. وبالتالي، إذا كانت c_k تكلفة إجراء مقابلة في الزيارة k فالتكلفة الكلية المتوقعة للقيام بـ i زيارة هي $n_o C(i)$، حيث،

$$C(i) = c_1 \sum w_{1j} p_j + c_2 \sum (w_{2j} - w_{1j}) p_j + \cdots + c_i \sum (w_{ij} - w_{i-1,j}) p_j$$

مثال

يبيّن الجدول (١٣ـ ٨) مجتمعًا بثلاثة صفوف. وقد اعتُمدت الـ p_j و w_{ij} لتمثل مسوحًا تجري فيها مقابلة بالغ نختاره عشوائيًا. وفي الزيارة الأولى أُخذت احتمالات الحصول على مقابلة w_{1j} في الصفوف الثلاثة على أنها 0.6 ، 0.3 و 0.1 . وفي الزيارات

الثانية وما بعدها نأخذ 0.9 ، 0.5 و 0.2 . كاحتمالات شرطية لمقابلة شخص، افتقدناه سابقًا. وقد وُضعت هذه الأرقام أعلى من الاحتمالات الموافقة في الزيارة الأولى كي تمثل تأثير الطريقة الذكية في الاستفسار التي يستخدمها المعاين.

جدول (١٣-٨) الخواص المميزة للصفوف الثلاثة

	الصف 1	الصف 2	الصف 3
p_j	0.45	0.50	0.05
w_{ij}	$0.6 + (0.4)[1 - (0.1)^{i-1}]$	$0.3 + (0.7)[1 - (0.5)^{i-1}]$	$0.1 + (0.9)[1 - (0.8)^{i-1}]$
I μ_j	55	50	45
II μ_j	60	50	40

جدول (١٣-٩) عدد المقابلات، التكاليف لكل مقابلة والانحيازات

عدد الزيارات المطلوبة	عدد المقابلات التي تمت	متوسط الكلفة لكل مقابلة	I الانحياز	II الانحياز
1	$0.425n_o$	100	+1.118	+2.235
2	$0.771n_o$	105	+0.711	+1.421
3	$0.882n_o$	108	+0.421	+0.842
4	$0.933n_o$	110	+0.266	+0.532
5	$0.960n_o$	114	+0.180	+0.360

والمفردة التي نقدّرها هي نسبة مئوية لثنائي حد قريبة من 50% وقد اعتُبرت مجموعتان من الـ μ_j(II,I) وللتبسيط أُخذت تباينات ما ضمن الصفوف $\sigma_j^2 = \mu_j(100 - \mu_j)$ مساوية جميعها لـ 2500 . أما التكاليف النسبية لكل مقابلة كاملة في الزيارات المتتالية فقد أُخذت من الجدول (١٣-٦).

ويبيّن الجدول (١٣-٩) (ا) العدد الكلي المتوقع للمقابلات التي حصلنا عليها في i من الزيارات، (ب) التكلفة المتوسطة لهذه الزيارات على أساس المقابلة الواحدة، (ج) الانحياز $(\bar{\mu}_i - \bar{\mu})$ في التقدير $\bar{y}$ تحت الفروض I و II حول الـ μ_j.

وفي II على سبيل المثال، نجد متوسط المجتمع الحقيقي $\bar{\mu}$ مساويًا لـ 54% . والمتوسط $\bar{\mu}_1$ الذي حصلنا عليه من الزيارات الأولى هو 56.235% مما يعطي

الانحياز المبيّن في الجدول ومقداره %2.235 . وتخفض سياسة تستدعي ثلاث زيارات هذا الانحياز بمقدار %0.842 .

وقد قورنت قيم $MSE(\bar{y})$ التي حصلنا عليها من نفقة مالية معطاة، وذلك من أجل سياسات مختلفة لتكرار الزيارة. وفي المقارنات الأولى نجد أن المبلغ المالي كاف لأخذ $n_0=500$ إذا قمنا بزيارة واحدة فقط. ومن الجدول (١٣-٩) نجد أن توقع عدد المقابلات التي حصلنا عليها في الزيارة الأولى $E(n_1)=(500)(0.425)=212.5$ وإذا قُمنا بزيارتين فيجب أن ينخفض هذا العدد المتوقّع إلى $E(n_2)=212.05/1.05=202.4$ ، كي نحافظ على التكلفة نفسها، والأمر مماثل من 3 ، 4 و 5 زيارات. وقد عُوِّضت هذه القيم لِـ $E(n_i)$ في المعادلة (13.10) لتعطي $V(\bar{y})$ و $MSE(\bar{y})$.

جدول (١٣-١٠) قيم $MSE(\bar{y})$ من أجل سياسات مختلفة لتكرار الزيارة وتكلّف المبلغ نفسه

عدد الزيارات المطلوبة	$n_o = 500$ (الزيارات الأولى فقط)			$n_o = 1000$		$n_o = 2000$	
	لا انحياز	I*	II*	I	II	I	II
1	**11.8**	13.0	16.9	7.1	10.9	4.2	8.0
2	12.4	12.9	14.6	6.7	8.3	3.6	5.2
3	12.7	**12.9**	13.6	**6.5**	7.1	3.4	3.9
4	13.0	13.1	**13.4**	6.6	**6.9**	**3.3**	3.6
5	13.5	13.5	13.8	6.8	6.9	3.4	**3.5**

* تمثل هذه مجتمعات بقدر أصغر (I) وأكبر (II) من الانحياز كما عرفناه في الجدول (١٣-٨).

ويمثل الجدول (١٣-١٠) الـ MSE الناتجة من أجل ثلاثة حجوم للنفقات موافقة لِـ n_o يساوي 500 ، 1000 و 2000 وذلك في حالة زيارة واحدة. وعندما يكون $n_o=500$ فإن قيم $MSE(\bar{y})$ معطاة أيضًا في حالة «اللاانحياز» التي يكون كل μ_i فيها مساويًا لِـ 50 . ويبيّن هذا العمود تأثير تكرار الزيارة عندما لا يكون هذا التكرار ضروريًا، باعتبار أن اقتصار المسح على زيارة واحدة لا ينتج انحيازًا.

والسياسات التي تعطي أقل MSE (متوسط خطأ تربيعي) مبيّنة بطباعة غامقة. لنعتبر أولًا أصغر حجم عيّنة $n_o=500$ وإذا لم يكن تكرار الزيارات ضروريًا، فإن سياسة

تتـطلب من الـزيارات عددًا يبلغ الأربـع لا تنتـج إلا زيادة متـواضعـة فقط في الـ MSE . وفي I ، وهي تنطوي على أصغر قدر من الانحياز، تنتج السياسات المختلفة حوالي الـدقـة نفسهـا، مع أن 3 هي المثلى. وفي II ، نجد أن تكرار الزيارة من 3 إلى 5 مرات شيء مُرض، وتعطي الزيارة الواحدة حوالي 25% زيادة في الـ MSE فوق القيمة الصغرى.

ومن أجل حجوم أكبر للعيّنة يزداد العدد الأمثل لتكرار الزيارات إلى 4 أو 5 ويُنتج استخدام زيارة بمفردها خسائر أكبر في الدقة.

وهذا بالطبع توضيح فقط. وتتجلى أهمية الطريقة في أنه عندما تتجمع معلومات حول التكاليف والإنجازات النسبية، يمكن رسم سياسة اقتصادية لأي نوع معيّن من المسوح الإحصائية.

(١٣-٦) كسر المعاينة الأمثل بين غير المستجيبين

بعد القيام بالمحاولة الأولى للوصول إلى الأشخاص في العيّنة، هناك أسلوب آخر يعود إلى Hansen و Hurwitz (1946) وهو أن نأخذ عيّنة جزئية عشوائية ممن لم نتصل بهم، ونقوم بجهود خاصة لمقابلة كل شخص في هذه العيّنة الجزئية. وقد طُوّرت أولاً لمسوح جرت فيهـا المحـاولـة الابتـدائية عن طريق الـبريد، ونأخذ عيّنة جزئية من الأشخاص الذين لم يعيدوا الاستبيان مستكملاً، ثم نتصل بهم بطريقة أكثر تكلفة فنقابلهم شخصيًا.

ويمكن اعتبـار هذه الطريقة كتطبيق لتقنية المعاينة المضاعفة مع التقسيم إلى طبقات، المقدمة في الفقرة (١٢-٢)، حيث نأخذ أولاً عيّنة عشوائية بسيطة من n' من الوحدات. ليكن n_1' عدد الوحدات في العيّنة التي تقدّم البيان المطلوب، و n_2' العدد في الـزمـرة غير المستجيبة. وبجهود مكثفة نحصل فيما بعد على عيّنة جزئية عشوائية $n_2 = \nu_2 n_2'$ من الـ n_2' ويستخـدم Hansen و Hurwitz الـرمز $\nu_2 = 1/k$ بحيث يكون $n_2 = n_2'/k$

وفي إطار الفقرة (١٢-٢) توجد طبقتان. الطبقة الأولى وتتألف من أولئك الذين يستجيبـون للمحاولة الأولى، ويشكلون عيّنة مقاسة حجمها $n_1 = n_1'$ أي أن $\nu_1 = 1$

وتتألف الطبقة 2 من أولئك الذين قد يستجيبون للمحاولة الثانية حيث $n_2 = n_2'/k$ وتكلفة أخذ العيّنة هي ،

$$c_0 n' + c_1 n_1' + \frac{c_2 n_2'}{k} \tag{13.12}$$

حيث المقادير c هي التكاليف لكل وحدة : c_0 تكلفة القيام بالمحاولة الأولى ، c_1 تكلفة معالجة نتائج المحاولة الأولى ، و c_2 تكلفة الحصول على المعلومات في الطبقة الثانية ومعالجتها . وإذا كانت W_1 ، W_2 نسبتي المجتمع في الطبقتين فالتكلفة المتوقعة هي ،

$$C = c_0 n' + c_1 W_1 n' + \frac{c_2 W_2 n'}{k} \tag{13.13}$$

وكتقدير لـ $\bar{Y}$ نأخذ ،

$$\bar{y}' = w_1 \bar{y}_1 + w_2 \bar{y}_2 = \frac{(n_1' \bar{y}_1 + n_2' \bar{y}_2)}{n'} \tag{13.14}$$

حيث $\bar{y}_1$ و $\bar{y}_2$ متوسطا العيّنتين اللتين حجماهما $n_1 = n_1'$ و $n_2 = n_2'/k$. ومن النظرية (١٢-١) ، سيكون التقدير $\bar{y}'$ غير منحاز إذا حصلنا على إجابة من جميع أفراد العيّنة الجزئية العشوائية المختارة والتي حجمها $n_2 = n_2'/k$.

ومن العلاقة (١٢-٣) ، يكون تباين $\bar{y}'$ ،

$$\begin{aligned} V(\bar{y}') &= \left(\frac{1}{n'} - \frac{1}{N}\right) S^2 + \frac{W_2 S_2^2}{n'} \left(\frac{1}{\nu_2} - 1\right) \\ &= \left(\frac{1}{n'} - \frac{1}{N}\right) S^2 + \frac{(k-1) W_2 S_2^2}{n'} \end{aligned} \tag{13.15}$$

وعندئذ نختار الكميتين n' و k بحيث تجعلان الجداء $C(V + S^2/N)$ أصغر ما يمكن .

ومن (13.15) و (13.13) نجد ،

$$V + S^2/N = \frac{(S^2 - W_2 S_2^2)}{n'} + \frac{k W_2 S_2^2}{n'} \tag{13.16}$$

$$C = (c_0 + c_1 W_1) n' + \frac{c_2 W_2 n'}{k} \tag{13.17}$$

ومن متراجحة كوشي، تكون k المثلى،

$$k_{opt} = \sqrt{\frac{c_2(S^2 - W_2 S_2^2)}{S_2^2(c_0 + c_1 W_1)}} \qquad (13.18)$$

ويمكن اختيار الحجم الابتدائي للعيّنة n' بحيث يجعل C أصغر ما يمكن في حالـة V محدد، أو يجعـل V أصغـر ما يمكن في حالـة C محددة، وذلـك بحلّ (13.16) و (13.17) من أجل n' . إذا كان V محددًا فنجد،

$$n'_{opt} = \frac{N[S^2 + (k-1) W_2 S_2^2]}{(NV + S^2)} \qquad (13.19)$$

حيث V القيمة المحددة لتباين تقدير متوسط المجتمع.

وتتطلب الحلول معرفة W_2 ، وإن كان يمكن، في الغالب، تقدير W_2 من خبرة سابقة. وبالإضافة إلى S^2 ، الذي يجب تقدير قيمته سلفًا في أي مسألة «تحديد حجم عيّنة»، فإن الحلول تتضمن أيضًا S_2^2 ، التباين في طبقة غير المستجيبين. وقد يكون التنبؤ بقيمة S_2^2 أصعب؛ ومن المحتمل أنها سوف لا تساوي قيمة S^2 بالذات. وعلى سبيل المثال، في مسوح جرت بالبريد لمعظم أنواع النشاط الاقتصادي، كان المستجيبون يميلون إلى أن يكونوا من أصحاب الفعاليات الأكبر، وتكون تباينات ما بين الوحدات أكبر مما هي بالنسبة لغير المستجيبين.

وإذا لم يكن W_2 معــروفـًا تمامًا فيمكن اللجـوء إلى تقـريب مُرض بحسـاب قيمة n'_{opt} مؤقتًا من (13.18) و (13.19) مع مدى مفروض من قيم W_2 يقع بين 0 وحد أعلى أمـين. ونـتبـنّـى أكـبر قيم n'_{opt} في هذه الـسلسلة من القيم كحجم ابتـدائي للعيّنـة n'. وعنـد تلقّي الإجـابات من المسح بالبريد تكون قيمة n_2' معروفة. وفي استهدافنا لقيمة k التي سنستخدمها في هذه الطريقة، نستخدم التباين $v_c(\bar{y}')$ الشرطي على قيم معروفة لـ n_2 و n' . ويمكن الحصول على هذا التباين من (12.4) و (12.7) على الشكل،

$$V_c(\bar{y}') = S^2\left(\frac{1}{n'} - \frac{1}{N}\right) + \frac{(k-1) n_2' S_2^2}{n'^2} \qquad (13.20)$$

وتُحل المعادلة (13.20) لإيجاد قيمة k التي تعطي التباين الشرطي المرغوب. وعادة تكون تكلفة هذه الطريقة أعلى بصورة طفيفة فقط من التكلفة المثلى الموافقة لـ W_2 معروف.

مثال

هذا المثال مكثف من بحث لـ Hansen و Hurwitz (1946) . وقد أُخذت العيّنة الأولى بواسطة البريد وقد توقّع الباحث أن يكون معدل الاستجابة مساويًا 50% . والدقة المرغوبة هي تلك التي كنا سنحصل عليها من عيّنة عشوائية بسيطة حجمها 1000 فيما لو لم يوجد أي عدم استجابة. وتكلفة البريد لكل استبيان هي 10 سنتات، وتكلفة معالجة نتائج استبيان مستكمل هي 40 سنتًا. ويكلّف القيام بمقابلة شخصية 4.10 $.

كم استبيانًا ينبغي إرساله؟ وما هي النسبة المئوية من غير المستجيبين الذين تجب مقابلتهم؟

وبدلالة دالة التكلفة (13.12) تكون تكاليف الوحدة بالدولار كما يلي:

c_0 = تكلفة المحاولة الأولى = 0.1

c_1 = تكلفة معالجة المعلومات من مستجيب = 0.4

c_2 = تكلفة الحصول على معلومات من لامستجيب ومعالجتها = 4.5

ويمكن إيجاد القيم المثلى لـ n' و k من (13.19) و (13.18) وإذا افترضنا تساوي التباينين S^2 و S_2^2 وأن N كبير بكفاية، فعندئذ،

$$k_{opt} = \sqrt{\frac{c_2(1-W_2)}{c_0+c_1W_1}} = \sqrt{\frac{(4.5)(0.5)}{0.1+(0.4)(0.5)}} = \sqrt{7.5} = 2.739$$

$$n'_{opt} = \frac{S^2[1+(k-1)W_2]}{V} = 1000\{1+(1.739)(0.5)\}$$

$$= 1870$$

لاحظ أننا وضعنا $S^2/V = 1000$ أو $V = S^2/1000$ ، باعتبار أن هذا هو ما سيكونه تباين متوسط عيّنة حجمها 1000 لو أنها أُخذت بدون وجود أي عدم استجابة.

وبالتالي، يجب إيداع 1870 استبيانًا في البريد. ومن بين الـ 935 التي لا تعود، نقابل عيّنة جزئية عشوائية من 935/2.739 أو 341 والتكلفة 2095 $.

وكما أشار Durbin (1954) فمن غير المحتمل أن تُظهر المعاينة الجزئية ربحًا ملحوظًا ما لم تكن c_2 كبيرة بالنسبة إلى $(c_0+c_1W_1)$ والكميتان قابلتان للمقارنة، باعتبار أن $(c_0+c_1W_1)$ هي التكلفة المتوقعة لكل وحدة عند القيام بالمحاولة الأولى ومعالجة نتائجها. ولـ c_2 المعنى نفسه بالنسبة للمحاولة الثانية. ومن المعادلات يمكن إثبات أن نسبة تكلفة الحصول على V محددة مع $k=1$ (بدون معاينة جزئية) إلى التكلفة الصغرى مع قيمة مثلى لـ k هي،

$$\frac{S^2(c_0+c_1W_1+c_2W_2)}{[\sqrt{(S^2-W_2S_2^{\,2})(c_0+c_1W_1)}+\sqrt{c_2}W_2S_2]^2}=\frac{c_0+c_1W_1+c_2W_2}{[\sqrt{W_1(c_o+c_1W_1)}+\sqrt{c_2}W_2]^2}$$

هذا إذا كان S^2 و $S_2^{\,2}$ متساويين تقريبًا. وإذا كانت r نسبة c_2 إلى $(c_0+c_1W_1)$ تصبح نسبة التكلفتين

$$\frac{1+rW_2}{(\sqrt{W_1}+\sqrt{r}W_2)^2}$$

وعلى سبيل المثال، نجد في حالة $W_1=0.5$ أن نسبة التكلفتين 1.029 إذا كان $r=4$، و 1.146 إذا كان $r=10$، و 1.228 إذا كان $r=16$. إلا أنه إذا كان S^2 أكبر بكثير من $S_2^{\,2}$ فهناك الكثير مما يمكن كسبه من معاينة جزئية.

ومع معاينة طبقية تكون القيم المثلى لـ n'_h و k_h في كل طبقة بمفردها معقدة إلى حد كبير، وكتقريب جيد، نقدّر أولاً، مستخدمين طرق الفقرتين (٥-٥) و(٥-٩)، حجوم العيّنات n_{oh} التي سنحتاجها من الطبقات فيما لو لم يوجد أي عدم استجابة. ومن (13.19) نجد الآن، في حالة $W_2=0$ أن:

$$n_o=\frac{NS^2}{NV+S^2} \tag{13.21}$$

وبالتالي يمكن كتابة (13.19) على الشكل،

$$n'_{opt}=n_o\left[1+\frac{(k-1)W_2S_2^{\,2}}{S^2}\right] \tag{13.22}$$

وتعطي هذه المعادلة عند تطبيقها بصورة منفصلة على كل طبقة، قيمة مُثلى تقريبية لِـ n_h'. وقد وُجدت قيم k_h بتطبيق (13.18) في كل طبقة.

ويمكن استخدام هذه الأساليب في تقديرات الانحدار والتقديرات النسبة. إذ نضع في التقديرات النسبة S_d^2 و S_{2d}^2 بدلاً من S^2 و S_2^2 حيث $d_i = y_i - Rx_i$. وفي تقديرات الانحدار نضع $S^2(1-\rho^2)$ بدلاً من S^2 و $S_2^2(1-\rho^2)$ بدلاً من S_2^2.

(١٣-٧) تعديلات من أجل الانحياز دون تكرار الزيارة

اقترح Hartley (1946) طريقة لامعة لتخفيض الانحيازات التي تظهر في نتائج الزيارة الأولى وطوّرها Politz و Simmons (1949, 1950) و Simmons (1954). لنفرض أن جميع الزيارات قد تمّت خلال المساء وذلك في أمسيات الأسبوع الست. وسُئل المستجيب عما إذا كان موجودًا في البيت في مثل وقت المقابلة خلال الأمسيات الخمس السابقة من الأسبوع. وإذا أفاد المستجيب بأنه كان في المنزل في t أمسية من بين 5 أمسيات تؤخذ النسبة $(t+1)/6$ كتقدير لتكرار وجوده في المنزل، π ، خلال ساعات المقابلة.

وتُخزن نتائج الزيارة الأولى مصنفة في 6 زمر وفقًا لقيمة t ، (0,1,2,3,4,5). وفي الزمرة t ليكن n_t عدد المقابلات التي تمت و $\bar{y}_t$ متوسط المفردة. فتقدير Politz-Simmons لمتوسط المجتمع μ هو،

$$\bar{y}_{PS} = \frac{\sum_{t=0}^{5} 6n_t\bar{y}_t/(t+1)}{\sum_{t=0}^{5} 6n_t/(t+1)} \qquad (13.23)$$

ويعترف هذا الأسلوب بأن نتائج الزيارة الأولى قد رُجّحت بدون مبرر بأشخاص يتواجدون في البيت معظم الوقت. باعتبار أن الشخص الذي يتواجد في البيت نسبة π من الوقت له، في المتوسط، فرصة نسبية للظهور في العيّنة مقدارها π ، وينبغي لإجابته أن تتلقى ترجيحة مقدارها $\frac{1}{\pi}$. وقد استخدمت الكمية $6/(t+1)$ كتقدير لِـ $\frac{1}{\pi}$. وهكذا يكون $\bar{y}_{PS}$ أقل انحيازًا من متوسط العيّنة من الزيارة الأولى $\bar{y}$ ، إلا أن

تباينه أكبر، لأننا استخدمنا متوسطًا مرجحًا بدلاً من متوسط غير مرجح مع ترجيحات مقدّرة تقديرًا.

وعند تقدير متوسط وتباين $\bar{y}_{PS}$ نستخدم رموز الفقرة (١٣-٥). ويُقسم المجتمع إلى صفوف، حيث الأشخاص في الصف j هم الذين يتواجدون في المنزل نسبة π_j من الوقت. ونلاحظ أن الزمرة t (أي الأشخاص الذين يتواجدون t أمسية من بين الأمسيات الخمس السابقة) ستتضمن أشخاصًا من صفوف مختلفة. ليكن $\bar{y}_{jt}, n_{jt}$ العدد الموافق لمفردة ومتوسط هذه المفردة لأولئك الموجودين في الصف j والزمرة t. فيمكن عندئذ كتابة $\bar{y}_{PS}$ كما يلي،

$$\bar{y}_{PS} = \frac{\sum\sum 6n_{jt}\bar{y}_{jt}/(t+1)}{\sum\sum 6n_{jt}/(t+1)} = \frac{N}{D} \quad \text{(مثلاً)} \qquad (13.24)$$

وهو من نوع التقدير النسبة. وفي العيّنات الكبيرة يكون متوسطه مساويًا لـ $E(N)/E(D)$ تقريبًا.

وإذا كان n_o الحجم الابتدائي للعيّنة (المستجيبون مضافًا إليهم غير الموجودين في المنزل) و n_j عدد من قوبلوا من الصف j، فنضع الفرضيات التالية،

$$\text{(i)} \quad p_j\pi_j \text{ تقدير ذي الحدين لـ } \quad \frac{n_j}{n_o}$$

$$\text{(ii)} \quad E(n_{jt}|n_j) = n_j\frac{5!}{t!(5-t)!}\pi_j^t(1-\pi_j)^{5-t}$$

$$\text{(iii)} \quad E(\bar{y}_{jt}) = \mu_j, \quad \text{لأي } j \text{ و } t$$

والفرض (ii) موضع تساؤل لأنه، وبدون الدخول في التفاصيل، يفترض أن الناس يقدمون إجابات صحيحة عن أوقات تواجدهم في البيت.

وفي حالة j معطى، مستخدمين الفرض (ii)،

$$E\sum_{t=1}^{5} n_{jt}\left(\frac{6}{t+1}\right) = n_j \sum_{t=1}^{5}\left(\frac{6}{t+1}\right)\frac{5!}{t!(5-t)!}\pi_j^t(1-\pi_j)^{5-t} \qquad (13.25)$$

$$= \frac{n_j}{\pi_j}[1-(1-\pi_j)^6] \qquad (13.26)$$

وبالتالي، مستخدمين الفرض (i) نجد،

$$E(D)=\sum_{j=1}^{r}\frac{E(n_j)}{\pi_j}[1-(1-\pi_j)^6]=n_o\sum_{j=1}^{r}p_j[1-(1-\pi_j)^6] \qquad (13.27)$$

وفضلاً عن ذلك، وباعتبار أن $E(\bar{y}_{jt})=\mu_j$ لأي j و t فإن هذا يعطي النتيجة،

$$E(\bar{y}_{PS})=\bar{\mu}_{PS}\doteq\frac{\sum_{j=1}^{r}p_j\mu_j[1-(1-\pi_j)^6]}{\sum_{j=1}^{r}p_j[1-(1-\pi_j)^6]} \qquad (13.28)$$

وبما أن المتوسط الصحيح $\bar{\mu}=\sum p_j\mu_j$ فيبقى بعض الانحياز في $\bar{y}_{PS}$ ومن ناحية يمكننا القول بإن لهذا التقدير الانحياز نفسه الذي نجده في $\bar{y}_6$ ، أي متوسط العيّنة الذي تعطيه طريقة الزيارات المتكررة التي تستدعي تكرار الزيارة 6 مرات إذا اقتضت الضرورة. وقد برهنّا في الفقرة (١٣-٥)، معادلة (13.9) أن طريقة الزيارات المتكررة، مع عدد كلي من الزيارات يساوي i تعطي تقديرًا غير منحاز لِـ $\bar{\mu}_{ji}=\sum w_{ij}p_j\mu_j/\sum w_{ij}p_j$ حيث w_{ij} احتمال أن تتمّ مقابلة شخص من الصف j ممن تضمهم العيّنة .والآن $w_{1j}=\pi_j$ وفي الزيارات اللاحقة، إذا بقي احتمال العثور في البيت على شخص لم نقابله سابقًا مساويًا π_j، فعندئذ،

$$w_{ij}=[1-(1-\pi_j)^i]$$

أي أن $\bar{\mu}_{PS}=\bar{\mu}_6$ إلا إنه في طريقة الزيارات المتكررة، قد يكون احتمال مقابلة في زيارة لاحقة أكبر من π_j كنتيجة لمعلومات حصل عليها المعاين في زيارته الأولى أو في إحدى زياراته السابقة. وفي هذه الحالة يكون لطريقة الزيارات المتكررة انحياز أقل بعد 6 زيارات.

وتباين $\bar{y}_{PS}$ معقد إلى حد بعيد. ومع التقريب المعتاد للتقدير النسبة، يمكن التعبير عنه، متّبعين Deming (1953) على الشكل التالي،

$$V(\bar{y}_{PS})\doteq\frac{1}{n_oU}\{\sum\pi_jp_jB_j[\sigma_j^2+(\mu_j-\bar{\mu}_{PS})^2]$$
$$+(n_o-1)\sum(\pi_jp_j)^2(B_j-A_j^2)(\mu_j-\bar{\mu}_{PS})^2\}$$

حيث،

$$U = 1 - \sum p_j(1-\pi_j)^6$$

$$A_j = \frac{1}{\pi_j}[1-(1-\pi_j)^6]$$

$$B_j = \sum_{t=0}^{5}\left[\frac{6}{(1+t)}\right]^2 \frac{5!}{t!(5-t)!}\pi_j^t(1-\pi_j)^{5-t}$$

ومع أنه يصعب تثمين هذه العبارة دون تطبيقها على مجتمعات محددة، يمكننا القيام بتعليقين. فإذا لم تختلف الـ μ_j اختلافًا كبيرًا، أي إذا كان الانحياز من الزيارات الأولى معتدلاً، فالحد المسيطر هو الحد الأول

$$\frac{1}{n_o U}\sum \pi_j p_j B_j \sigma_j^2$$

وتميل هذه العبارة إلى أن تكون 25% إلى 35% أعلى من تباين المتوسط غير المرجّح للزيارات الأولى. ويتضمن $V(\bar{y}_{PS})$ أيضًا حدًّا لا يتناقص مع زيادة n_o ويصبح مهمًّا في عيّنات كبيرة جدًّا.

ويمكن التلخيص بالقول إن مقارنات قمت بها مع Deming (1953) و Durbin (1954) وتناولت مجتمعات اصطُنعت بالمحاكاة، تقترح أن هذه الطريقة تُظهر تفوقًا ملحوظًا بالمقارنة مع تكرار الزيارات، عندما تكون الانحيازات من الزيارات الأولى كبيرة والعيّنة كبيرة. وعلى أي حال، تكون التخفيضات في متوسط مربعات الخطأ MSE من أجل النفقات نفسها صغيرة، ما لم تكن كلفة تكرار الزيارات أكبر بكثير مما هو مفترض هنا. وتتمتع طريقة Politz-Simmons بميزة توفير الوقت. أما الأخطاء وعدم الكمال في قيم t فتعتبر من السيئات. ويمكن تطبيق الطريقة أيضًا، كما اقترح Simmons (1954) على حد سواء مع تعدد الزيارات.

وقد اقتُرحت عدة طرق أخرى لتخفيف الانحياز الناشىء عن «ليس موجودًا في البيت». وتنطبق طريقة Bartholomew (1961) على مسح يتطلب زيارتين. ويفترض، من أجل أولئك الذين لم يُشاهدوا في الزيارة الأولى، أنه يمكن للمعاين جعل احتمالات

مشاهدتهم في الزيارة الثانية متساوية تقريبًا، وذلك من خلال الاستفسار بعناية. وإذا كان الأمر كذلك فإن الأشخاص الـ n_2 الذين قوبلوا في الزيارة الثانية هم عيّنة عشوائية جزئية من الـ(n_o-n_1) شخصًا الذين افتقدنا وجودهم في الزيارة الأولى. وبالتالي يكون $[n_1\bar{y}_1+(n_o-n_1)\bar{y}_2]/n_o$ تقديرًا غير منحاز لمتوسط العيّنة الابتدائية المستهدفة. وقد أدّت الطريقة أداء حسنًا في بعض المسوح الإحصائية البريطانية التي طبقها فيها Bartholomew . ويقترح Kish و Hess (1959a) في حالة مسوح إحصائية متكررة، أن عدم الاستجابات من مسوح حديثة يمكن أن تخدم كبديل لعدم الاستجابات في مسح إحصائي قيد التنفيذ. وحيثما يُظهر الانحياز من زيارات مبكرة نموذجًا نمطيًا، كما في الجدول (١٣-١)، يقدّم Hendricks (1949) الخطوط العريضة لطرق توسّع استقرائي لتقدير متوسط النتائج التي يمكن أن يعطيها اللامستجيبون.

(١٣-٨) نموذج رياضي لأخطاء القياس

يمكننا أن نتصور ذهنيًا إمكانية القيام بعدد كبير من التكرارات المستقلة للقياس من الوحدة i ليكن $y_{i\alpha}$ القياس الذي نحصل عليه من التكرار α فعندئذ،

$$y_{i\alpha}=\mu_i+e_{i\alpha} \tag{13.29}$$

حيث

$\mu_1=$ القيمة الصحيحة

$e_{i\alpha}=$ خطأ القياس.

وتستدعي فكرة «القيمة الصحيحة» قليلاً من النقاش. ففي بعض المفردات تكون الفكرة بسيطة ومحددة تمامًا. وعلى سبيل المثال، في جرد للبضائع مأخوذ بطريقة المعاينة، يمكن أن تكون القيمة الصحيحة عدد قشاطات المراوح الموجودة على الرف في الساعة الثانية عشرة ظهرًا من يوم معيّن. وفي بعض الحالات يمكن تعريف القيمة الصحيحة بصورة عملياتية. فمثلاً يمكن تعريف ضغط الدم الانبساطي الصحيح لشخص في وقت محدد بأنه القيمة التي نحصل عليها من استخدام جهاز معياري معين تحت شروط موصوفة بعناية. إلا إنه يمكن التحقق من أن جهازنا المعياري هو نفسه خاضع لأخطاء القياس. ويمكننا أن نتوقع، مع مرور الزمن، أنه سيجري تطوير جهاز

أكثر دقة. ومن أجل مفردات أخرى، على سبيل المثال، شيء عن موقف مستخدم تجاه رب العمـل، أو عن مشاعر شخص حول مقدرته على معالجة مشاكله اليومية، فلا يستطيع أحد الادّعاء بأن لديه طريقة مُرضية لقياس «القيمة الصحيحة». ومع ذلك فالفكرة مفيدة حتى في حالات كهذه.

ومع تكرار القياسات من الوحدة نفسها سيتبع الخطأ $e_{i\alpha}$ توزيع تكرار ما. ومن أجل الوحدة i، لنفرض أن متوسط $e_{i\alpha}$ هو β_i وتباينه σ_i^2. ويمثل الحد β_i انحيازًا في القياسات. وستعتمد مقادير β_i و σ_i^2، بالطبع، على طبيعة المفردة التي نقيسها وعلى جهـاز القياس. وقد تعتمد أيضا على الكثير من العوامل الأخرى. وفي المجتمعات البشرية يمكن للمناخ الاقتصادي والسياسي السائد ولمقدار ونوع الشعبية التي تلقّاها المسح الإحصائي أن تؤثر في المستجيبين للاستبيان الإحصائي.

والخطوة التالية هي أن ندرس كيفية تغير أخطاء القياس عندما نتحرك من وحدة إلى أخرى. فقد تنشأ تعقيدات متنوعة.

ومن أجل مركبة الانحياز β_i، قد يكون هناك انحياز ثابت، مثلاً، $E(\beta_i)=\beta$، يؤثر في جميع وحدات المجتمع. وسيكون هناك أيضًا مركبة $(\beta_i-\beta)$ تتبع توزيعًا تكراريًا ما فوق المجتمع. ويمكن أن تكون هذه المركبة مرتبطة بالقيمة الصحيحة μ_i، وعلى سبيل المثـال، يمكن لأداة القياس، وبـاستمـرار، أن تقـدّر بالنقصـان قيمًا مرتفعـة لـ μ_i وتقدّر بالزيادة قيمًا منخفضة.

وقد يوجد ارتباط بين قيم $e_{i\alpha}$ في وحدات مختلفة من العيّنة نفسها. وأبسط مثال هو «انحياز المعاين». وتوجد أحيانًا فروق هائلة في متوسط قيم $y_{i\alpha}$ كما يحصل عليها معـاينـون مختلفـون يقـومـون بمعاينة أجزاء قابلة للمقارنة من المجتمع نفسه (انظر Lienau ,1941، Mahalanobis ,1946 و Barr ,1957).

وقـد ظهرت تأثيرات مماثلة عند جني عيّنات من محصول زراعي بوساطة فِرق مختلفة، وعند القيام بتحليلات كيميائية أو بيولوجية في مخابر مختلفة. والعامل الإنساني ليس السبب الوحيد لوجود ارتباط بين الوحدات التي تُقاس في الوقت نفسه تقريبًا. ويتـأثـر العديد من عمليات القياس بالطقس؛ والبعض يستخدم مواد خام تختلف نوعيتهـا من دُفعـة إلى دُفعة. وعند تقدير سعر البيع الحالي لمنازل شيّدت منذ بضع

سنين، يشير Hansen ، Hurwitz و Bershad (1961) إلى أنه إذا بيعت بعض المنازل في العيّنة حديثًا فإن أسعارها تُرسي مستوى يُرشد المعاين وصاحب المنزل في تخصيص قيم لمنازل لم يجرِ بيعها منذ عديد من السنين. وفي الحقيقة، قد يعتمد السعر المتوسط المسجّل للعيّنة، على ترتيب ظهور المنازل المباعة حديثًا في العيّنة.

ولكي نعالج هذه الارتباطات ضمن العيّنة في أعم أشكالها فإننا نحتاج إلى نموذج أكثر تعقيدًا من ذلك المقدّم هنا. وبصورة خاصة، ربما كان من الواجب أن تشير الرموز $e_{i\alpha}$ و β_i إلى إمكانية اعتماد قيمها على وحدات أخرى موجودة ضمن العيّنة. وعلى أي حال، يمكن تمثيل أنواع الارتباط، التي يُعتقد أنها الأكثر شيوعًا في التطبيقات العملية، بالنموذج الحالي أو بتعميمات بسيطة له.

ومركبات خطأ القياس ملخصة في الجدول (١٣-١١). وقد لاحظنا أيضًا أن قيم β_i و $d_{i\alpha}$ في وحدات مختلفة من العيّنة نفسها يمكن أن ترتبط ببعضها البعض، حيث $d_{i\alpha} = e_{i\alpha} - \beta_i$.

جدول (١٣-١١) مركبات خطأ القياس في الوحدة i

طبيعة المركبة	الرمز
انحياز ثابت فوق جميع الوحدات	β
مركبة انحياز متغيرة تتبع عندما يتغير i توزيعًا تكراريًا ما متوسطه الصفر ويمكن أن تكون على ارتباط بالقيمة الصحيحة μ_i	$\beta_i - \beta$
مركبة خطأ متأرجحة تتبع عند تغير α مع بقاء i ثابتًا، توزيعًا تكراريا ما بمتوسط صفر وتباين σ_i^2	$d_{i\alpha} = e_{i\alpha} - \beta_i$

وقد طوّر Hansen وآخرون (1951) ، Sukhatme و Seth (1952) Hansen, ، Hurwitz و Bershad (1961) و Hansen ، Hurwitz و Pritzker (1965) ، نماذج مماثلة بصورة عامة للنموذج السابق. وأعطى Fellegi (1964) نموذجًا أكثر شمولاً يتضمن العديد من الارتباطات التي قد توجد بين مصادر مختلفة لخطأ القياس في الحالات العملية.

وفي بحثيهم في 1961 و 1965 عبّر Hansen وآخرون عن نموذجنا بمصطلحات تختلف اختلافًا طفيفًا بما في ذلك إضافة دليل G إلى جميع المتغيرات للتذكير بأن الأخطاء وحجومها يمكن أن تعتمد على الشروط العامة (G) للمسح الإحصائي. ويعبرون عن نتائجهم بدلالة الـ $d_{i\alpha}$ وبدلالة:

$$\mu_i' = E(y_{i\alpha}|i) = \mu_i + \beta_i \tag{13.30}$$

والكمية μ_i' (التي يرمزون إليها بـ Y_i أو P_i في حالة نسبة) هي من حيث مفهومها الذهني المتوسط الحاصل على العديد من تكرارات عملية القياس في الوحدة i. وبالتالي،

$$d_{i\alpha} = e_{i\alpha} - \beta_i = y_{i\alpha} - \mu_i' \tag{13.31}$$

ويدعون $d_{i\alpha}$ انحراف الاستجابة في الوحدة i، بينما دعوناه نحن مركبة تأرجح خطأ القياس. وهكذا يكون،

$$y_{i\alpha} - \mu = d_{i\alpha} + (\mu_i' - \mu') + (\mu' - \mu) \tag{13.32}$$

حيث μ' متوسط المجتمع للمقادير μ_i'.
وبأخذ المتوسط فوق العيّنة، نجد،

$$\bar{y}_\alpha - \mu = \bar{d}_\alpha + (\bar{\mu}' - \mu') + (\mu' - \mu) \tag{13.33}$$

وهذا يعطي العلاقة،

$$\text{MSE}(\bar{y}_\alpha) = V(\bar{d}_\alpha) + V(\bar{\mu}') + (\mu' - \mu)^2 + 2\,\text{cov}\,(\bar{d}_\alpha, \bar{\mu}') \tag{13.34}$$

ومن أجل متوسط العيّنة $\bar{y}_\alpha$، تدعى الحدود على اليمين، على الترتيب، تباين الاستجابة، تباين المعاينة، مربع الانحياز الإجمالي، وضعف التغاير بين متوسط انحراف الاستجابة في العيّنة وخطأ المعاينة. وبما أن $E(d_{i\alpha}|i) = 0$ تحت نموذجنا، فإن التغاير ينعدم تحت تكرار عملية القياس في المجموعة نفسها من وحدات العيّنة ووفق الترتيب نفسه. وقد لا ينعدم تحت تكرارات فوق عيّنات مختلفة أو ترتيبات مختلفة إذا تأثرت الـ $d_{i\alpha}$ بالوحدات الأخرى في العيّنة. ومع أننا سنتجاهل هذا الحد هنا توخيًا للبساطة، فقد بيّن Fellegi (1964) في دراسة كندية (نسبة إلى كندا) أنه يمكن أن يؤدي إلى انحياز مهم في بعض طرق تقدير مركبات تباين الاستجابة $V(\bar{d}_\alpha)$.

وباستخدام أسلوب Cornfield كما وصفناه في الفقرة (٢-٩)، أعطى Koch (1973) تفكيكًا عامًّا لمتوسط مربعات خطأ التقدير في مسوح عيّنة متعددة المتغيرات، مع تطبيقات تتناول متوسطات صفوف جزئية.

(١٣-٩) تأثيرات انحياز ثابت

لنفرض أن القياسات y_i في جميع الوحدات تخضع لانحياز ثابت β كميته معروفة. فعندئذ يخضع متوسط عيّنة عشوائية بسيطة $\bar{y}$ أيضًا للانحياز β نفسه. ويُلغى الانحياز في تقدير تباين الخطأ الموافق لمتوسط العيّنة، باعتبار أن هذا التقدير مستخلص من مجموع مربعات الحدود $(y_i - \bar{y})^2$، وبالتالي فإن الحساب العادي من بيان العيّنة لحدود الثقة الموافقة لِـ $\bar{Y}$ لا تأخذ أي اعتبار للانحياز. وتصح النتائج نفسها في معاينة عشوائية طبقية.

وتبقى الحالة نفسها، في الأساس، من أجل التقدير النسبة وتقدير الانحدار. ولنعتبر تقدير الانحدار،

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$$

حيث يمكن أن تخضع الـ y_i والـ x_i إلى انحيازين ثابتين β_y و β_x على الترتيب. وبما أن تقدير المربعات الدنيا b يبقى بدون تغيير، وأن الانحياز β_x يُلغى من الحد $(\bar{X} - \bar{x})$ فينتج أن $\bar{y}_{lr}$ خاضع للانحياز β_y. ومن السهل التحقق من أن تقدير العيّنة لِـ $V(\bar{y}_{lr})$ لا يحوي أية مساهمة تعود إلى مركبتي الانحياز.

وفي حالة التقدير النسبة،

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}} \bar{X}$$

يكون الانحياز أيضًا β_y كتقريب أول، باعتبار أن $E(\bar{X} / \bar{x})$ هو الواحد تقريبًا في العيّنات الكبيرة، حتى لو كانت الـ x_i خاضعة لانحياز ثابت. وفي العيّنات الكبيرة يكون تقدير التباين،

$$v(\bar{y}_R) = \frac{(N-n)}{Nn} \frac{\sum (y_i - \hat{R} x_i)^2}{n-1} \tag{13.35}$$

حرًّا تقريبًا من الانحياز، وذلك باعتباره تقديرًا لِـ

$$E(\bar{y}_R - \bar{Y})^2$$

أي أنه تقدير للتباين حول المتوسط المنحاز $\bar{Y}$.

ونلخص بالقول إن الانحياز الثابت يمرّ دون أن يكشفه البيان الإحصائي للعيّنة. وكما رأينا (الفقرة ١ ـ ٨)، فإن الـ 95% احتمالات ثقة لا تتأثر تقريبًا إذا كانت نسبة β_y إلى الخطأ المعياري لتقدير المتوسط أقل من 0.1 ، ولكن كلما ازدادت النسبة عن هذه القيمة، تصبح حسابات حدود الثقة مضلّلة. وتبقى تقديرات التغير، من فترة زمنية إلى أخرى، أو من طبقة إلى أخرى، غير منحازة، وذلك شريطة بقاء الانحياز ثابتًا على الدوام.

(١٣-١٠) تأثيرات الأخطاء غير المرتبطة ضمن العيّنة

في هذه الفقرة نعطي القليل من النتائج حول $V(\bar{y}_\alpha)$ و $v(\bar{y}_\alpha)$ ، من أجل معاينة عشوائية بسيطة، وذلك في أبسط الحالات حيث تكون أخطاء القياسات غير مرتبطة ضمن العيّنة. ويمكن تطبيق هذه الحالة في مسوح إحصائية مأخوذة من سجلات، في استبيانات إحصائية تُملأ ذاتيًا من قبل صاحب العلاقة (كما في المسوح البريدية) حيث لا يستشير أعضاء في العيّنة نفسها بعضهم بعضًا، وفي بعض المجتمعات غير الحيّة التي يكون القياس فيها موضوعيًا. ولا يمكن القيام بفرض عدم الارتباط دون رويّة: قد يدخل الارتباط ضمن العيّنة، على سَبيل المثال، من خلال عملية المقابلة، الطباعة، الترميز، أو تحويل المعلومات الإحصائية إلى حاسب آلي، إذا كان الشخص يعالج عددًا من العناصر في العيّنة.

ومن (13.34) نجد، من أجل التباين،

$$V(\bar{y}_\alpha) = V(\bar{d}_\alpha) + V(\bar{\mu}') \qquad (13.36)$$

ولإيجاد التباينات نحسب أولاً المتوسط فوق تكرارات عملية القياس على عيّنة محددة، ومن ثَمّ فوق عيّنات عشوائية بسيطة مختلفة. والآن $V(d_{i\alpha}) = \overline{\sigma_i^2}$، والأخطاء غير مرتبطة في وحدات مختلفة ضمن العيّنة. وهكذا نجد من أجل معاينة عشوائية بسيطة،

حيث يرمز لعيّنة محددة، ما يلي:

$$E(\bar{d}_\alpha{}^2|S) = \frac{1}{n^2}\sum_i^n \sigma_i{}^2 \tag{13.37}$$

وعندما نأخذ المتوسط فوق جميع العيّنات العشوائية البسيطة نجد بالتالي:

$$V(\bar{y}_\alpha) = \frac{1}{nN}\sum_i^N \sigma_i{}^2 + \frac{1-f}{n}\frac{\sum_i^N(\mu_i{}'-\mu)^2}{N-1} \tag{13.38}$$

$$= \frac{1}{n}\sigma_d{}^2 + \frac{(1-f)}{n}S_{\mu'}{}^2 \tag{13.39}$$

حيث يرمز $\sigma_d{}^2$ لمتوسط تباينات أخطاء القياس مأخوذًا فوق المجتمع. ووفقًا لمصطلحات Hansen وآخرون يكون σ_d^2/n تباين الإجابة لمتوسط العيّنة $\bar{y}_\alpha$.

ومع أخطاء غير مرتبطة يمكن تطبيق النموذج نفسه لتقدير نسبة المجتمع P (Hansen ، Hurwitz و Bershad ,1961) . ومن أجل أي وحدة، لتكن القيمة الصحيحة μ_i مساوية للواحد إذا كانت الوحدة i في الصفه C وصفرًا فيما عدا ذلك، فوقوع أخطاء في القياس يتضمن أن الوحدات تُصنّف أحيانًا تصنيفًا خاطئًا. ومن أجل الوحدة i تكون القيمة المسجلة $y_{i\alpha}$ هي الواحد أحيانًا والصفر أحيانًا أخرى. لنرمز بـ P_i لنسبة القياسات في الوحدة i التي يكون فيها $y_{i\alpha}=1$ فعندئذ، وفي حالة i معطى، يكون $y_{i\alpha}$ متغيرًا ثنائيًا عند تكرار القياس، بمتوسط $\mu_i{}' = P_i$ بينما تباين $d_{i\alpha}$ هو P_iQ_i . وهكذا إذا كان $p_\alpha = \bar{y}_\alpha$ تقدير العيّنة، فإن (13.38) تصبح،

$$V(p_\alpha) = V(\bar{y}_\alpha) = \frac{1}{nN}\sum^N P_iQ_i + \frac{(1-f)}{n}\frac{\sum^N(P_i-P)^2}{N-1} \tag{13.40}$$

$$< \frac{1}{n(N-1)}\left(\sum^N P_i - \sum^N P_i^2 + \sum^N P_i^2 - NP^2\right)$$

$$= \frac{N}{n(N-1)}PQ \tag{13.41}$$

حيث

$$P = \frac{\sum^N P_i}{N}$$

وكما تبينّ (13.41) نجد أن لمجموع تباين الإجابة وتباين المعاينة حدًّا أعلى $NPQ/n(N\text{-}1)$ وهذا الحد الأعلى هو أيضًا تباين متوسط العيّنة للقياسات الصحيحة في الوحدات، إذا تجاهلنا الـ ت م م ولم يكن هناك انحياز إجمالي. ذلك لأن القياس الصحيح μ_i في أي وحدة سيكون عندئذ متغيراً ثنائيًا بمتوسط P ، وبالتالي،

$$\frac{S_\mu^{\ 2}}{n}=\frac{N}{n(N-1)}PQ\geq V(p_\alpha) \qquad (13.42)$$

وتصحّ هذه النتيجة المحيّرة إلى حد كبير لأن: (i) تباين المعاينة الداخل في (13.39) و (13.40) هو تباين$|(\mu_i+\beta_i)=\mu_i{}'$ و (ii) عند تقدير نسبة يكون الارتباط بين μ_i و β_i سالبًا دومًا. فعندما يكون $\mu_i=1$ يكون $\beta_i\leq 0$ باعتبار أن $P_i=(\mu_i+\beta_i)\leq 1$، وبصورة مشابهة عندما $\mu_i=0$، $\beta_i\geq 0$. وهكذا يكون الحد،

$$\frac{S_{\mu'}^{\ 2}}{n}$$

في (13.39) أقل دائمًا من

$$\frac{S_\mu^{\ 2}}{n}$$

بما يساوي تقريبًا تباين الإجابة لـ $\bar{y}_\alpha$ فقط.

ومع أخطاء غير مرتبطة نجد نتيجة مفيدة هي أن العلاقة من أجل $v(\bar{y}_\alpha)$ في معاينة عشوائية بسيطة تبقى غير منحازة إذا تجاهلنا الـ ت م م. ومن نتيجة النظرية (٢-٤)، تصبح هذه العلاقة، عند تطويرها تحت الفرض بعدم وجود أخطاء قياسات، على الشكل،

$$v(\bar{y}_\alpha)=\frac{1-f}{n}s^2=\frac{1-f}{n}\frac{\sum^n (y_{i\alpha}-\bar{y}_\alpha)^2}{(n-1)} \qquad (13.43)$$

والآن،

$$y_{i\alpha}-\bar{y}_\alpha=(d_{i\alpha}-\bar{d}_\alpha)+(\mu_i{}'-\bar{\mu}') \qquad (13.44)$$

وبالتربيع ثم أخذ المتوسط، أولاً فوق القياسات المتكررة، ثم فوق الاختيارات المتكررة للعيّنة، نجد،

$$Ev(\bar{y}_\alpha) = \frac{1-f}{n}\sigma_d{}^2 + \frac{1-f}{n}S_{\mu'}{}^2 \qquad (13.45)$$

بينما نجد من (13.39) ،

$$V(\bar{y}_\alpha) = \frac{1}{n}\sigma_d{}^2 + \frac{(1-f)}{n}S_{\mu'}{}^2 \qquad (13.46)$$

وهكذا يكون $Ev(\bar{y}_\alpha) = V(\bar{y}_\alpha)$ إذا أمكن إهمال f .

وبالطريقة نفسها يمكن البرهان على أن العلاقات في الفصل السابق الخاصة بتقديرات العيّنة لتباينات خطأ المعاينة تبقى صحيحة في معاينة طبقية أو معاينة متعددة المراحل، وكذلك فإن علاقات العيّنة الكبيرة قابلة للتطبيق على تقدير الانحدار والتقدير النسبة، شريطة أن تكون أخطاء القياس في $y_{i\alpha}$ و $x_{i\alpha}$ غير مرتبطة ضمن العيّنة وأن يكون الـتمم مهملاً .

(١٣-١١) تأثيرات الارتباط ضمن العيّنة بين أخطاء القياسات

لنفرض وجود ارتباط بين بعض أو جميع قيم $d_{i\alpha}$ من أجل وحدات في العيّنة نفسها . فيمكن كتابة الحد $\bar{d}_\alpha{}^2$ على الشكل،

$$\bar{d}_\alpha{}^2 = \frac{1}{n^2}\left(\sum_i^n d_{i\alpha}{}^2 + 2\sum_{i<j}^n d_{i\alpha}d_{j\alpha}\right) \qquad (13.47)$$

وبالتالي، إذا أخذنا المتوسط فوق قياسات متكررة، وعيّنات عشوائية بسيطة، نجد،

$$V(\bar{d}_\alpha) = E(\bar{d}_\alpha{}^2) = \frac{1}{n}\sigma_d{}^2 + \frac{2n(n-1)}{2n^2}E(d_{i\alpha}\, d_{j\alpha}) \qquad (13.48)$$

حيث أخذنا متوسط الجداءات فوق جميع أزواج الوحدات في العيّنة نفسها . وقياسًا على المعاينة العنقودية، يمكن تعريف متوسط معامل الارتباط ضمن العيّنة ρ_w بالمعادلة،

$$E(d_{i\alpha}\, d_{j\alpha}) = \rho_w \sigma_d{}^2 \qquad (13.49)$$

وهذا يعطي من (13.48) ،

$$V(\bar{d}_\alpha) = \frac{\sigma_d{}^2}{n}[1 + (n-1)\rho_w] \qquad (13.50)$$

وقد سمّى Hansen ، Hurwitz و Bershad (1961) $V(\bar{d}_\alpha)$ تباين الإجابة الكلي باعتبار أن له تأثيره في متوسط العيّنة. وتُدعى مركبته σ_d^2/n تباين الإجابة البسيط، بينما يدعى الحدّ $(n-1)\rho_w\sigma_d^2/n$ امركبة الارتباط لتباين الإجابة الكلي.

ومن (13.34) نجد، مفترضين $\text{cov}(\bar{d}_\alpha, \bar{\mu}') = 0$ ،

$$V(\bar{y}_\alpha) = \frac{\sigma_d^2}{n}[1+(n-1)\rho_w] + \frac{1-f}{n}S_{\mu'}^2 \qquad (13.51)$$

وبالطريقة نفسـها وُجد أن القيمة المتوسطة لِـ $v(\bar{y}_\alpha)$ العادي في (13.34) هي،

$$Ev(\bar{y}_\alpha) = \frac{(1-f)}{n}[\sigma_d^2(1-\rho_w) + S_{\mu'}^2] \qquad (13.52)$$

وبما أنه من المحتمل أن يكون ρ_w موجبًا في العديد من أنواع خطأ القياس، فالعلاقة المُتعارف عليها لِـ $v(\bar{y}_\alpha)$ تشكل عادة تقديرًا بالنقصان في هذه الحالة، وتجعل تقدير العيّنة يبدو وكأنه أكثر دقة مما هو عليه فعلاً. وهذا صحيح حتى عندما يكون الـ ت م م مهملاً.

وربما كان المثال الأكثر تكرارًا لارتباط ما ضمن العيّنة بين أخطاء القياسات هو ارتباط ما ضمن المعاين المذكور سابقًا، وبصورة خاصة حول مسائل تنطوي على آراء وأحكـام. ولنفرض أن $n = mk$ وأن كلاً من المعاينين الـ k يحصل على بيان إحصائي من m مستجيبًـا. وإذا استـطعنا الافتراض بأنه لا يوجد ارتباط بين أخطاء القياس لمعاينين مختلفين فعندئذ،

$$V(\bar{d}_\alpha) = \frac{\sigma_d^2}{n}[1+(m-1)\rho_w] \qquad (13.53)$$

حيث ρ_w متوسط معامل ارتباط ما ضمن المعاين.

ويمكن حتى لِـ ρ_w صغير أن يقدم إسهامًا رئيسًا في $V(\bar{y}_\alpha)$ باعتبارها مضروبة على وجه التقريب بحجم حصة المعاين m . وقد يوجد أيضًا بعض الارتباط بين أخطاء القياس لمعاينين مختلفين (مثلاً، إذا كانوا قد تلقوا تدريباتهم أو وُجهوا من قِبل المشرف نفسه).

ولا يمثل هذا النموذج إلا أبسط نوع من ارتباط ما ضمن العيّنة. وفي معاينة طبقية، على سبيل المثال، يمكن لمن يقوم بالترميز أن يفرّغ النتائج من عدة طبقات،

ومن خلال التباس معينّ في التعليمات يمكن أن يُدخل أخطاء مرتبطة تعمّ الطبقات جميعها. ويمكن تكييف النموذج الرياضي بحيث ينطبق على حالات من هذا النوع.

(١٣-١٢) خلاصة تأثيرات أخطاء القياسات

بدلالة النموذج يمكن أن يكون المتوسط $\bar{y}$ لعيّنة عشوائية بسيطة غير منحاز وأن يكون تباينه $S_\mu{}^2/n$ (متجاهلين الـ ت م م)، هذا إذا كانت القياسات جميعها دقيقة. وكنتيجة لأنواع أخطاء القياس التي نوقشت هنا، يمكن أن يخضع المتوسط لانحياز قدره β، ويكون متوسط مربعات خطئه،

$$\text{MSE}(\bar{y}_\alpha) = \frac{1}{n}\{S_{\mu'}{}^2 + \sigma_d{}^2[1 + (n-1)\rho_w]\} + \beta^2 \qquad (13.54)$$

حيث $\mu_i' = \mu_i + \beta_i$.

وتتضمن العلاقة (13.54) حدّين $S_{\mu'}{}^2/n$ و $\sigma_d{}^2(1-\rho_w)/n$ يتناقصان كلما تناقص $\frac{1}{n}$. ويبدو للوهلة الأولى أن الحدين $\rho_w\sigma_d{}^2$ وβ^2 مستقلان عن n. وقد يكون هذا تبسيطًا مبالغًا فيه. وأي تغيّر كبير في حجم العيّنة يمكن أن يستدعي تغيرًا في الطرق الميدانية للقياس، وقد يؤثر هذا في ρ_w وβ^2. وعلى أي حال، ينبغي أن يتغير هذان الحدان تغيرًا بطيئًا نسبيًا مع n، إذا كان هناك أي تغير. وهكذا فمن المحتمل أن يسيطر هذان الحدان، في العيّنات الكبيرة، على متوسط مربعات الخطأ MSE مما يجعل تباين المعاينة العادي مضللاً وغير ذي شأن كدليل على الدقة الحقيقة للنتائج.

(١٣-١٣) دراسة أخطاء القياس

كُرّس جزء من البحث في مجال تطبيقات المعاينة في السنوات الأخيرة لدراسة أخطاء القياس. وكانت الأهداف هي اكتشاف المركبات التي تُسهم إسهامًا كبيرًا في الـ MSE وإيجاد طرق لتخفيض هذه المساهمات. ونعرض في هذه الفقرة والفقرات التالية بعضًا من الطرق الرئيسة. ويتضح لنا مما سبق أن التقدم سيكون بطيئًا ومكلفًا. وأحد الأسباب هو أن أخطاء القياس، كما ذكرنا سابقًا، تعتمد اعتمادًا حميمًا على كل من المفردات وطريقة القياس. ونادرًا ما يمكننا الافتراض بأن النتائج التي نتوصل إليها حول أخطاء القياس في مسح ما يمكن تطبيقها على مسوح أخرى.

ومن وجهـة النـظر النمـوذجية فإن أفضل طريقة لدراسة أخطاء القياس هي الحصـول على القيم الصحيحـة μ_i . وفي الممارسة العملية تقتصر هذه الطريقة على مفردات تتوافر من أجلها طريقة ممكنة للوصول إلى μ_i وتبقى المشاكل هنا مشاكل نفقة وتنفيذ. وهناك أمثلة يعطيها Belloc (1951) الذي قام بمقارنة المعلومات المتعلقة بدخول المستشفى والمـأخوذة من زيارات للمنازل مع سجلات المستشفى الخاصة بالشخص المعني، و Gray (1955) الذي قارن إفادات المستخدمين المتعلقة بالإجازات المَرَضية مع سجـلات دائـرة الذاتية. وتكون تحقيقات من هذا النوع ـ وتدعى أحيانًا «تحقيقات السجل» ـ ممكنة مع مفردات مثل العمر، المهنة، عدد سنوات المدرسة، والثمن المدفوع لسيارة. وإحـدى الصعـوبـات هي أن السجلات لا تتضمن أحيانًا تلاؤمًا تامًّا مع الشخص الذي جرت مقابلته.

وعنـد افتقـاد طريقة لتحديد القيمة الصحيحة يكون أحد البدائل هو إعادة القياس بطريقـة مستقلة نعتبرها أكثر دقة. وقد كلّف Kish و Lansing (1954) خبراء تخمين محترفين بتقدير أثمان بيع المنازل التي كان أصحابها قد أفادوا من حينهم بثمنها. وفي مسوح تتعلق بالمرض قورنت إجابات المستجيبين إما بسجلات أطبائهم أو بنتائج فحص طبي كامل [Dunham, Sagen و Trussell, 1959 Simmons و 1959 Elinson] وليس من السهل تفسير نتائج مقارنات كهذه بدلالة النموذج، طالما أن الأداة المتفوقة هي نفسها خاضعة لأخطاء القياس، إلا أن المقارنات ستشير على الأقل إلى المفردات التي تتفق من أجلها، وبصورة جيدة، الأداتان الروتينية والمتفوقة، وإلى المفردات التي لا يتحقق من أجلها مثل هذا الاتفاق.

وإحدى الطرق ذات الهدف المماثل، هي إعادة مقابلة عيّنة جزئية من المستجيبين في مسوح زيارات المنازل بوساطة معاينين أكثر خبرة واستبيان أكثر تفصيلاً وأدق سبراً. وبعد إعادة المقابلة يناقش الخبير مع المستجيب أي تباين بين الأجوبة الأصلية والأجوبة المُعادة، والهدف هنا هو تحديد الجواب الأكثر دقة وسبب التباين. وقد تُكتسب الكثير من المعلومـات المفيدة. وفي تقديمه لبعض وسائل قياس أخطاء الإجابات، يناقش

Madow (1965) استخدام المعاينة المضاعفة، مع مقدّر فرقي من الشكل $[\bar{y}' - (\bar{y} - \hat{\bar{\mu}})]$ كوسيلة لتخفيض انحياز الإجابة. و $\hat{\bar{\mu}}$ هنا هو متوسط القياسات غير المنحازة أو الأقل انحيازًا التي تمّت في العيّنة الجزئية، بينما $\bar{y}'$ و $\bar{y}$ هما متوسطا القياسات الأصلية في العيّنة والعيّنة الجزئية.

ومن وقت لآخر، يمكن إجراء مقارنات إجمالية بين نتائج مَسحين مختلفين. ومن أجل عدد من المفردات يمكن مقارنة نتائج مكتب التعداد الإحصائي في الولايات المتحدة مع النتائج المأخوذة في الوقت نفسه التي يعطيها المسح الراهن للمجتمع (Current Population Survey). وبما أن المسح يُعتبر أكثر دقة، وبصورة خاصة في مفردات صعبة القياس، فيمكن القيام بتقديرات تقريبية لِـ β انحياز القياس في معلومات مكتب التعداد (Hurwitz, Hansen و Bershad ,1961) وقد ناقش Stephan و Mc Carthy (1958) عددًا من المقارنات بين نتائج عيّنات بالحصّة وعيّنات احتمالية.

وندرس في الفقرات التالية بعض الطرق المصمّمة لإنتاج تقديرات كمية لمركبات التباين الكلي (تباين الإجابة مضافًا إليه تباين المعاينة) لتقدير عيّنة.

(١٣-١٤) إعادة قياس عيّنات جزئية

يتركّز الاهتمام في الدراسات الحديثة على: (ا) تباين الإجابة الكلي والحجوم النسبية لتباين الإجابة البسيط ولمركبة الارتباط كمُساهمين فيه، و(ب) الحجوم النسبية لتباين الإجابة الكلي وتباين المعاينة. وإحدى الطرق الشائعة من أجل (ا) هي اختيار عيّنة جزئية من العاملين في القياس (معاينين، عاملين في الترميز، الخ.) ثم إعادة قياس الحصص المخصصة لهم بواسطة آخرين يُفترض أنهم في مستوى المهارة نفسه. لنفرض أن حجم العيّنة الجزئية mk حيث m حجم حصة معاين و k عدد المعاينين الذين اختيروا للعيّنة الأصلية.

ليكن y_{i1}, y_{i2} القياسين من الوحدة i من حصّة معاين. إذا صحّت (13.13) فعندئذ،

$$y_{i\alpha} = d_{i\alpha} + \mu_i' \qquad (13.55)$$

وبالتالي إذا أخذنا المتوسط فوق الحصة نجد،

$$\frac{E\sum_i^m (y_{i1} - y_{i2})^2}{2m} = \frac{\sigma_{d1}^2 + \sigma_{d2}^2}{2} - \text{cov}\,(d_{i1}\,d_{i2}) \qquad (13.56)$$

وهذه العبارة تقدّر σ_{d1}^2 تباين الإجابة البسيط للمعاين 1 في المسح، في حالة صحة الشرطين (A) وهما: لا ارتباط بين أخطاء الاستجابة d_{i1}، d_{i2} في الوحدة نفسها، و $\sigma_{d1}^2 = \sigma_{d2}^2$ وتنطبق المعادلة (13.65) على زوج بمفرده من المعاينتين، ويؤخذ متوسطها فوق العيّنة الجزئية التي تتضمن k من أزواج المعاينين.

ويمكن أن تصحّ الشروط A عندما يكون القياس عملية ترميز، والمعاينون الذين يقومون بالترميز ذوو مهارات متماثلة ومدرّبين من قِبل مشرفين مختلفين، ولا يرى أحد المعاينين عمل الآخر. وفي حالة الزيارات يجب أن نتوقع $\text{cov}(d_{i1}d_{i2})$ موجبًا لأن بعض المستجيبين يكرر أول إجابة خاطئة من ذاكرته. وفي هذه الحالة يكون $\sum^m (y_{i1} - y_{i2})^2/2m$ تقديرًا بالنقصان لِـ σ_{d1}^2 وهو أيضًا تقدير بالنقصان لِـ σ_{d1}^2 إذا كان المعاين الثاني أمهر من المعاين الأول، كما يبيّن Hurwitz ، Hansen و Pritzker (1965) من أجل قياس من النوع (0,1) . وفضلاً عن ذلك، فقد وُجد أن $\sum (y_{i1} - y_{i2})^2/2m$ ينخفض إذا أُعطي المعاين الثاني الإجابات التي حصل عليها المعاين الأول، حتى ولو طُلب منه ألا ينظر إليها حتى تتم المقابلة المكررة (1973, Koons) وتوضح هذه التعقيدات لماذا تكون الدراسة الواقعية لأخطاء القياس صعبة.

ومن أجل تباين الإجابة الكلي يكون التقدير المناسب من (13.55) لزوج بمفرده من المعاينين هو $(\bar{y}_{.1} - \bar{y}_{.2})^2/2$ حيث $\bar{y}_{.1}, \bar{y}_{.2}$ هما متوسطا الـ m قياسًا للمعاين الأول والـ m قياسًا للمعاين الثاني، على الترتيب. وتحت الشروط (A) نجد،

$$\frac{E(\bar{y}_{.1} - \bar{y}_{.2})^2}{2} = \frac{\sigma_d^2}{m}[1 + (m-1)\rho_w] \qquad (13.57)$$

حيث ρ_w الارتباط بين أخطاء الإجابة من وحدات مختلفة يتعهدها المعاين نفسه. وتقدم المعادلة (13.57) درجة واحدة من الحرية فقط، إلا أنه يؤخذ متوسطها فوق الأزواج

الـ k من المعـاينـين. وبأخذ تقدير لِـ σ_d^2 من (13.56) يمكن تقدير الحجوم النسبية لتباين الإجابة البسيط ومركبة الارتباط. وفي مسوح الزيارات، وُجد أن مركبة الارتباط أكبر عادة بكثير من تباين الإجابة البسيط، باستثناء ما يتعلق بمفردات أساسية مثل العمر، الجنس، الحالة الاجتماعية (1964, Fellegi) وإلى حد ما، كان هذا هو سبب استخدام إحصاء 1970 في الولايات المتحدة للتعداد الذاتي البريدي استخدامًا واسعًا (Hansen و Waksberg, 1970).

وإذا كانت (13.55) صحيحة، فيمكن أيضًا دراسة نسبة تباين الإجابة البسيط إلى تباين المعاينة مطبقة على العيّنة التي تتضمنها حصّة معاين. ذلك لأنه تحت الشروط A،

$$E\sum_{\alpha}^{2}\sum_{i}^{m}(y_{i\alpha}-\bar{y}_{.\alpha})^2/2(m-1)=\sigma_d^2(1-\rho_w)+S_{\mu'}^2 \qquad (13.58)$$

وقد دعا Pritzker و Hanson (1962) النسبة

$$I=\frac{\sigma_d^2}{(\sigma_d^2+S_{\mu'}^2)} \qquad (13.58)'$$

دليل عدم الاتساق. وهو مشابه للكمية $(1-\phi)$ حيث ϕ معامل الموثوقية المستخدم في دراسة أخطاء القياس في علم النفس. وإذا كان ρ_w مهملاً، فيمكن تقدير I من نسبة (13.56) إلى (13.58).

ومع متغير (0,1) إذا وضعنا $n=1$ فتبيّن المعادلات (13.40) و (13.41) في الفقرة (١٣ـ١٠) أنه في حالة قياس لوحدة بمفردها يكون مجموع تباين الإجابة البسيط وتباين المعاينة $V(y_{1\alpha})\doteq PQ$ حيث $P=\sum P_i/N$، وبفرض f مهملاً. ويمكن تلخيص مجموعة القيم المشتركة (y_{i1}, y_{i2}) التي يحصل عليها معاينان في الجدول التكراري 2×2 التالي:

عدد الاستجابات

		المعاين الثاني 1	0	المجموع
المعاين الأول	1	a	b	$a+b$
	0	c	d	$c+d$
		$a+c$	$b+d$	m

وهكـذا يكـون b عدد الوحدات التي يسجل فيها المعاين الأول 1 ، ويسجل الثاني 0 . وتحت الشرطين A ، نجد،

$$\hat{\sigma}_d^{\,2}=\frac{\sum_{i}^{m}(y_{i1}-y_{i2})^2}{2m}=(b+c)/2m \qquad (13.59)$$

وأحد الاختيارات لتقدير I من العيّنة الجزئية هو،

$$\hat{I}=\frac{\hat{\sigma}_d^{\,2}}{\hat{P}\hat{Q}}=\frac{m(b+c)}{(a+b)(c+d)+(a+c)(b+d)} \qquad (13.60)$$

حيث يؤخـذ متـوسط تقـديرات الـ PQ من مجموعتي القياسات. وقد نشر Pritzker و Hanson (1962) ، Fellegi (1964) ، Koons (1973) تقديرات للدليل في حالة مفردات كل من تعـداد إحصـائي ومسـح عيّنة. وهذه التقديرات مفيدة لمقارنة عدم الموثوقية النسبية للقياسات في حالة مفردات مختلفة وفي تعدادات إحصائية متتالية، أو لمقارنة طرق مختلفـة للقياس، مما يزودنـا ببعض القـدرة على تثمين طرق جديدة للقياس. وبالطبع فإن تفسير هذه المقارنات تظلله غالبًا الشكوك فيما إذا كانت الشروط A مطبقة أم لا. ويعطي Fellegi (1964) تقديرات أكثر تعقيدًا قائمة على شروط أكثر واقعية.

(١٣-١٥) عيّنات جزئية متداخلة

كان Mahalanobis (1946) قد اقترح هذه الطريقة المفيدة بصورة خاصة في دراسة أخطاء مرتبطة. ولتقديمها في أبسط العبارات نقول إن عيّنة عشوائية من n وحدة يجري تقسيمها عشوائيًا إلى k من العيّنات الجزئية، وتتضمن كل عيّنة جزئية $m=n/k$ وحدة. ويجري تخطيط العمل الميداني ومعالجة معلومات العيّنة بحيث لا يوجد ارتباط بين أخطاء القياس لأي وحـدتين من عيّنتين جزئيتين مختلفتين. وعلى سبيل المثال، لنفرض أن الارتباط الذي نضطر للتعامل معه ينشأ فقط عن انحيازات المعاينين. إذا خُصّصت عيّنة جزئية مختلفة لكل من المعاينين الـ k ولم يكن هناك ارتباط بين أخطاء القياس عند معاينين مختلفين فلدينا مثال عن الطريقة.

ومـع النمـوذج الـرياضي نفسـه يكـون من المـريح أن نرمز للوحدات بدليل مضاعف. ليكن،

$$y_{ij\alpha} = \mu'_{ij} + d_{ij\alpha} \tag{13.61}$$

حيث يرمز i للعيّنة الجزئية (معاين) و j للعضو ضمن العيّنة الجزئية. والـ ت م م مهمل.

وبـما أن العيّنـة الجزئية i هي عيّنة جزئية عشوائية، فإنها نفسها عيّنة عشوائية بسيطة حجمها m، وهكذا يكون تباين متوسطها، استنادًا إلى (13.51)،

$$V(\bar{y}_{i\alpha}) = \frac{1}{m}\{S_{\mu'}^2 + \sigma_d^2[1 + (m-1)\rho_w]\} \tag{13.62}$$

حيث ρ_w الارتباط بين الـ $d_{ij\alpha}$ التي يحصل عليها المعاين نفسه.

وبما أن الأخطاء مستقلة في العيّنات الجزئية المختلفة، فنجد،

$$V(\bar{y}_{\alpha}) = \frac{1}{k} V(\bar{y}_{i\alpha}) = \frac{1}{n}\{S_{\mu'}^2 + \sigma_d^2[1 + (m-1)\rho_w]\} \tag{13.63}$$

ومن نتائج العيّنة يمكن حساب تحليل تباين الـ km من الملاحظات إلى مركبتي «ما بين المعاينين» (العيّنات الجزئية) بـ $(k\text{-}1)$ درجة من الحرية و«ما ضمن المعاينين» بـ $k(m\text{-}1)$ درجة من الحرية. ويمكن التحقق بسهولة من أن القيم المتوقعة لمتوسطات المـربعـات، حيث تؤخـذ المتـوسـطات فوق اختيارات المعـاينين والعيّنات العشوائية والعيّنات الجزئية، هي كما في الجدول (١٣-١٢).

جدول (١٣-١٢) توقعات متوسطات المربعات (على أساس وحدة بمفردها)

	df	ms	E(ms)
ما بين المعاينين	$k-1$	$s_b^2 = \dfrac{m\sum(\bar{y}_{i\alpha} - \bar{y}_{\alpha})^2}{k-1}$	$S_{\mu'}^2 + \sigma_d^2[1 + (m-1)\rho_w]$
(العينات الجزئية)	$k(m-1)$	$s_w^2 = \dfrac{\sum\sum(y_{ij\alpha} - \bar{y}_{i\alpha})^2}{k(m-1)}$	$S_{\mu'}^2 + \sigma_d^2(1 - \rho_w)$

ويتضمن الجدول (١٣-١٢) نتيجتين مهمتين. وبالمقارنة مع (13.63) نرى أن $V(\bar{y}_\alpha)$ تقدير غير منحاز لـ s_b^2/n متجاهلين الـ ت م م. وهكذا تقدم العيّنات الجزئية المتداخلة تقديرًا لـ $V(\bar{y}_\alpha)$ يأخذ بصورة مناسبة اعتبارًا لكل من تباين الإجابة البسيط ومركبة الارتباط.

ويمكننا التحليل أيضًا من تقدير مركبة الارتباط، باعتبار أن،

$$\frac{E(s_b^2 - s_w^2)}{m} = \rho_w \sigma_d^2 \qquad (13.64)$$

وبالتالي، تقدّرمقارنة $(m-1)\ (s_b^2 - s_w^2)/m$ مع s_b^2 المقدار النسبي لمساهمة مركبة ارتباط تباين الإجابة في التباين الكلي لِـ $\bar{y}_\alpha$. ومع القياسات التي تكون فيها مركبة الارتباط أكبر بكثير من تباين الإجابة البسيط، استُخدمت النسبة $(m-1)(s_b^2 - s_w^2)/ms_b^2$ كقياس بديل للمساهمة النسبية لتباين الإجابة الكلي في التباين الكلي لِـ $\bar{y}_\alpha$. ويقدم Tepping و Boland (1972) تقديرات لهذه النسبة من أجل مفردات في المسح الراهن للمجتمع (Current Population Servey) .

وعند تطبيق طريقة العيّنات الجزئية المتداخلة في عيّنة متعددة المراحل تغطي مساحة جغرافية واسعة. فإن الممارسة الأكثر شيوعًا هي أن يقيس زوج من المعاينين عيّنات جزئية مسحوبة من العناقيد الأصغر بين المراحل المتتالية. وبهذه الطريقة نحافظ على كون عدد الوحدات النهائي في حصّة معاين في المسح ضمن مستواه المعتاد، مع أن المعاين مضطر للتنقل فوق ضعف المساحة المعتادة. وفي عنقود بمفرده يقدم الجدول (١٣-١٢) درجة واحدة من الحرية لما بين المعاينين و $2(m\text{-}1)$ درجة من الحرية لما ضمن المعاينين: أُخذ معدّل متوسط المربعات الموافق فوق العناقيد الـ c التي اختيرت للدراسة. وتباين المعاينة الذي قيس ليس، بالطبع، إلا تباين ما ضمن المرحلة الأخيرة من المعاينة العنقودية. وكتذكير بهذه الحقيقة يُكتب حد تباين المعاينة في $E(ms)$ أحيانًا، على الشكل $S_{\mu'}^2(1-\rho_s)$ بدلاً من $S_{\mu'}^2$ حيث ρ_s الارتباط ضمن العنقود بين أخطاء المعاينة في وحدات جزئية مختلفة.

وقـد استُخـدمت طريقـة الغيّنـات الجزئية المتداخلة بهذا الشكل في الدراسة I لتبـاين الإجـابة التي قام بها مكتب الإحصاء في الولايات المتحدة (1968) ، والتي صُممت لتقـدير مركبـات الارتبـاط لتبـاينـات الإجـابـة الكلية لمفـردات من تعداد 1960 . وكـانت المساحات في هذه الدراسة عناقيد متراصّة من المنازل، أما العناقيد فمنتشرة فوق كامـل الـولايات المتحـدة الأمـريكية. وقـد شُكّلت عيّنتـان جزئيتـان متداخلتان في كل عنقود تضمنته العيّنة من العناقيد. وخُصصت كل عيّنة جزئيّة إلى معاين مختلف.

وفي نصف العناقيد كان للمعاينين قائدي طاقم مختلفين، وفي هذا النصف، افتُرض، وهو فرض يبدو منطقيًا، أن أخطاء الإجابة بالنسبة للمعاينين غير مرتبطة. وهكذا يكون $(s_b{}^2-s_w{}^2)/m$ تقديرًا لـ $\rho_w\sigma_d{}^2$ حيث يمثل $s_b{}^2$ الآن معدل متوسط المربعات بين المعاينين في العنقود نفسه. وفي النصف الآخر من العناقيد كان للمعاينين قائد الطاقم نفسه. والهدف هنا كان قياس المدى الذي يُنتج فيه تأثير قائد الطاقم تغاير بين $\bar{d}_{.1}$ و $\bar{d}_{.2}$ من أجل المعاينين في عنقود واحد. وإذا كان الأمر كذلك فإن $s_b{}^2$ في الجدول (١٣-١٢) يقدّر الآن،

$$S_{\mu'}{}^2(1-\rho_s)+\sigma_d{}^2[1+(m-1)\rho_w]-mE\operatorname{cov}(\bar{d}_{.1}\bar{d}_{.2})$$

و $(s_b{}^2-s_w{}^2)/m$ يقدّر،

$$\rho_w\sigma_d{}^2-E(\operatorname{cov}\bar{d}_{.1}\bar{d}_{.2}) \tag{13.65}$$

ومقارنة المجموعتين من قيم $(s_b{}^2-s_w{}^2)/m$ تكشف عن وجود «تأثير قائد الطاقم». وبما أن الفروق بين تقديرات تباينين مثل $s_b{}^2$ و $s_w{}^2$ غير مستقرة، فمن الضروري أخذ عدد كبير من العناقيد في نصفي مثل هذا النوع من الدراسة كي نقيس لأي درجة من الدقة «تأثير قائد الطاقم».

ويمكن تعميم طريقة العيّنات الجزئية المتداخلة إلى معاينة طبقية ومعاينة متعددة المراحل. وإذا كان الاهتمام الرئيس ينصبّ على تقدير غير منحاز لـ $V(\bar{y}_a)$ ويأخذ

اعتبارًا مناسبًا لتأثيرات أخطاء القياس، فكل ما نضطر إليه هو أن تتألف العيّنة من عدد من العيّنات الجزئية، من البنية نفسها التي نطمئن فيها إلى أن أخطاء القياسات في عيّنات جزئية مختلفة هي أخطاء مستقلة عن بعضها. وعلى وجه الدقة يتطلب هذا استخدام فُرق معاينة مختلفة، مشرفين مختلفين، ومفرغي بيانات مختلفين، في عيّنات جزئية مختلفة. وإذا كان $\bar{y}_{i\alpha}$ متوسط العيّنة الجزئية i فإن $\sum(\bar{y}_{i\alpha} - \bar{y}_{\alpha})^2/k(k-1)$ يشكل تقديرًا غير منحاز لـ $V(\bar{y}_{\alpha})$، بـ $(k\text{-}1)$ درجة من الحرية. وتصحّ هذه النتيجة لأنه يمكن اعتبار العيّنة الجزئية كوحدة معاينة مركبة بمفردها، وتكون العيّنة في الواقع عيّنة عشوائية بسيطة من هذه الوحدات المركبة، بأخطاء قياس غير مرتبطة بين وحدات مركبة مختلفة. وبالتالي تنطبق نتائج الفقرة (١٣-١٠). ويصف Deming (1960) عددًا كبيرًا من تطبيقات هذه الطريقة التي تدعى أحيانًا المعاينة المتكررة، والتي استخدمها Deming على نطاق واسع. ومن أجل مناقشات أخرى لفوائد هذه الطريقة، انظر Jones (1955) و Koop (1960) وتزداد تكاليف سفر المعاينين في طريقة العيّنات الجزئية المتداخلة، إلا أنه يمكن تلطيف ذلك إذا قسمنا العيّنة إلى طبقات مؤلفة من مناطق متراصة. وعلى سبيل المثال، يمكن أن تتضمن كل طبقة عيّنتين عشوائيتين مخصصتين لمعاينين مختلفين. ويُطلب من كل معاين أن يتنقّل فوق الطبقة بكاملها بدلاً من أن يتنقّل فوق نصف الطبقة فقط. وتقدم كل طبقة درجة واحدة من الحرية لتقدير $V(\bar{y}_{\alpha})$.

(١٣ - ١٦) تركيب التداخل وتكرار القياس

كما رأينا (فقرة ١٣-١٤)، يقدّم تكرار قياس الوحدات في حصّة معاين من قِبل معاين آخر من نوعية مماثلة تقديرات لتباين الإجابة البسيط وتباين الإجابة الكلي، هذا إذا انطبقت الشروط A، مع أنها يمكن أن تكون تقديرات بالنقصان في مسوح زيارات، إذا كانت أخطاء المستجيبين في المناسبتين مرتبطة إيجابًا. وتقدّم خطة التداخل (فقرة ١٣-١٥) تقديرات مركبة الارتباط لتباين الإجابة وإسهامها في التباين الكلي، تباين الإجابة زائدًا تباين المعاينة.

ويمكن تعلّم المزيد من طريقة حاذقة لتركيب التداخل والتكرار استخدمها

Fellegi (1964) في دراسة أخطاء الإجابة في تعداد السكان في كندا عام 1961 . وقد نُفّذت الدراسة في 134 منطقة تعداد (E.A.'s) وكل منها يتضمن حوالي 150 منزلاً هي حجم حصّة المعاين. وقد جُمعت مناطق التعداد المتجاورة في 67 زوجًا. وخُصّص معاينان لكل زوج، يقابل كل منهم نصفًا عشوائيًا من المنازل في الزوج. وهكذا يكون لكل عدّاد الحمل المعتاد نفسه من العمل، إلا أن المنازل الخاصة به تنتشر فوق ضعف المساحة. ثم يتبادل المعاينان حصتيهما، مما يعطي التركيب المرغوب للتداخل وتكرار القياس.

وإذا رمزنا بـ S_1 ، S_2 للعيّنتين الجزئيتين المتداخلتين وبـ I_1 ، I_2 للمعاينَين، فإن مقارنة (I_1S_1) مع (I_2S_1) أو (I_1S_2) مع (I_2S_2) تعطي تحليل «تكرار القياس»، بينما تعطي مقارنة (I_1S_1) مع (I_2S_2) أو (I_1S_2) مع (I_2S_1) تحليل «التداخل». وتقود هذه المقارنات إلى تقديرات لتباين الإجابة البسيط ولمركبة الارتباط ولتباين الإجابة الكلي ولدليل عدم الاتساق. وتباين المعاينة المعني هنا هو التباين بين المنازل ضمن أزواج مناطق التعداد. ويعطي تحليل أكثر اتساعًا قام به Fellegi تقديرات أيضًا لتغاير انحرافات الإجابة والمعاينة للمعاين نفسه، وقد أُهمل هذا الحدّ في النموذج المقدّم هنا، إلا أن Fellegi يبيّن أن هذا الحد يمكن أن يخلق انحيازات ذات شأن في تقديرات $\rho\sigma_d^2$.

ويعطي Bailar و Dalenius (1969) عرضًا موفقًا لنقاط القوة والضعف في أشكال مختلفة لأسلوبي تكرار القياس والتداخل. ويراجع Hansen وWaksberg (1970) البحث الذي قام به مكتب الإحصاء في الولايات المتحدة (U.S. Census Bureau) حول أخطاء القياس باعتبارها تؤثر في التعداد وفي بعض أهم مسوح العيّنة التي قام بها مكتب الإحصاء. وقد تضمنت التغيرات التي طرأت على تعداد عام 1970 : استخدامًا أكثر اتساعًا للتعداد الذاتي (مما يلغي انحيازات المعاينين) في تعداد بوساطة البريد، والمزيد من استخدام المعاينة كشيء متميز عن التعداد التام، واختيارًا مسبقًا للعيّنة بوساطة الحاسب، وذلك لتجنّب بعض الانحيازات التي اكتُشفت في اختيار المعاينين لعيّنة

المرحلة الأخيرة. والأخطاء المزعجة التي عُثر عليها في البيانات المتعلقة بالمهنة، الصناعة، نوعية السكن بالإضافة إلى مشاكل التذكّر فيما يتعلق ببيان المصروفات، لا تزال جميعها قيد دراسة مستمرة.

(١٣-١٧) أسئلة حساسة: إجابات معشّاة

تقع الحالة التي يُحتمل أن تقود إما إلى رفض إعطاء جواب أو إلى أجوبة مراوغة عندما يكون أحد الأسئلة في المسح حسّاسًا أو شخصيًا إلى حد بعيد (مثلاً، هل يعمل المستجيب بصورة منتظمة في سرق المعروضات في متجر؟ أو هل يستعمل بعض العقاقير؟) لندرس في البداية تقدير نسبة توزيع ثنائي - النسبة π_A من المستجيبين الذين ينتمون إلى صف معيّن A أو ارتكبوا فعلاً معينًا. وباستخدام حاذق لأداة عشوائية بيّن Warner (1965) أنه يمكن تقدير هذه النسبة دون أن يكشف المستجيب عن واقعه الشخصي بالنسبة إلى هذا السؤال. والهدف هو تشجيع الأجوبة الصحيحة تمامًا مع الاحتفاظ بالسرية التامة.

وتختار الأداة العشوائية، مثل سهم يدور أو صندوق يتضمن كرات حمراء وبيضاء، إحدى عبارتين أو سؤالين، كل منهما يتطلب إجابة نعم أو لا، كي يجري تقديمه إلى المستجيب. وبالنسبة لأي مستجيب لا يعلم المعاين السؤال الذي أُجيب عنه من بين السؤالين، إلا أنه يعلم الاحتمالات النسبية P و $(1-P)$ التي قُدّمت بها العبارتان. ويعتمد نجاح الطريقة، بالطبع، على قناعة المستجيب بأنه سوف لا يكشف باشتراكه واقعه الشخصي بالنسبة للمسألة الحساسة.

وفي مقترح Warner الأصلي كانت العبارتان:

«أنا عضو في الصف A». (وتقدَّم باحتمال P)

«أنا لست عضوًا من الصف A»

وفي عيّنة عشوائية من n مستجيبًا يسجّل المعاين تقديرًا ثنائيًا $\hat{\phi} = m/n$ لِـ ϕ نسبة الأجوبة «نعم». وإذا أُجيب على الأسئلة بصدق تام، تكون العلاقة بين ϕ و π_A في المجتمع كما يلي،

$$\phi = P\pi_A + (1-P)(1-\pi_A) = (2P-1)\pi_A + (1-P) \quad (13.66)$$

ومع P معروفة تقترح هذه العلاقة التقدير،

$$\hat{\pi}_{AW} = \frac{[\hat{\phi} - (1-P)]}{(2P-1)} \quad \left(P \neq \frac{1}{2}\right) \quad (13.67)$$

ويتمخض هذا التقدير عن كونه تقدير الإمكانية العظمى لِـ π_A (يرمز الدليل W إلى Warner) والتقدير غير منحاز بتباين هو:

$$V(\hat{\pi}_{AW}) = \frac{\phi(1-\phi)}{n(2P-1)^2} \quad (13.68)$$

وبكتابة $(1-\phi)$ على الشكل،

$$(1-\phi) = (2P-1)(1-\pi_A) + (1-P) \quad (13.69)$$

نجد بسهولة أن،

$$V(\hat{\pi}_{AW}) = \frac{\pi_A(1-\pi_A)}{n} + \frac{P(1-P)}{n(2P-1)^2} \quad (13.70)$$

والحد الأول في $V(\hat{\pi}_{AW})$ هو $V(\hat{\pi}_A)$، لو أن جميع المستجيبين أجابوا بصدق عن السؤال المتعلق بانتمائهم إلى الصف A . وباستثناء قيم π_A القريبة من $\frac{1}{2}$ و $P>0.85$ يكون الحد الثاني أكبر من الأول، وغالبًا أكبر بكثير. وهكذا تكون الطريقة غير دقيقة بصورة عامّة. ويمكن توقع هذا، باعتبار أن المعاين لا يعلم ما إذا كان الجواب نعم يتضمن الانتماء إلى الصف A أو العكس. وكما بيّن Warner فإن طريقته يمكن أن تعطي، على أي حال، متوسط مربعات خطأ، MSE أصغر مما كان يمكن أن يعطيه سؤال حسّاس مباشر، إذا أدى هذا الأخير إلى عدد كبير من حالات رفض الإجابة أو الإجابات الكاذبة.

(١٣ـ ١٨) السؤال الثاني الغريب

كبديل لطريقة Warner اقترح (Horvitz ، Simmons ، Shah و Simmons ، 1967) أن تعاون المستجيب يمكن أن يتحسّن إذا لم تكن العبارة الثانية حسّاسة بأي شكل من الأشكال، باعتبار أنه لا صلة لها بالأولى. مثلاً،

«ولدت في شهر أيار (مايو)» .

وتبقى العبارة الأولى بدون تغيير. وإذا أجاب الجميع بصدق تكون نسبة الأجوبة «نعم» في المجتمع ،

$$\phi = P\pi_A + (1-P)\pi_U \tag{13.71}$$

حيث π_u نسبة مواليد أيار (مايو) في المجتمع الذي نعاينه . وإذا كان معروفًا ، فإن التقدير الواضح (وهو تقدير إمكانية عظمى) لِـ π_A يكون ،

$$\hat{\pi}_{AU} = \frac{[\hat{\phi} - (1-P)\pi_U]}{P} \tag{13.72}$$

بتباين هو،

$$V(\hat{\pi}_{AU}) = \frac{\phi(1-\phi)}{nP^2} \tag{13.73}$$

ويقترح Morton (Greenberg وآخرون ,1969 صفحة 532) كيف يمكننا دائمًا تحقيق الحالة التي يكون فيها π_u معروفًا. نأخذ صندوقًا يتضمن كرات حمراء وبيضاء وزرقاء بنسب معروفة P_1 ، P_2 و P_3 . وينتج سحب كرة حمراء العبارة الحساسة. أما سحب كرة بيضاء أو زرقاء فيُنتج العبارة: «لون هذه الكرة أبيض». وبذلك يكون $\pi_U = P_2/(P_2+P_3)$.

وقد بيّن Dowling و Shachtman (1975) أن $V(\hat{\pi}_{AU}) < V(\hat{\pi}_{AW})$ من أجل جميع القيم لِـ π_A ، π_U شريطة أن يتجاوز P الثلث تقريبًا. (تباين $\hat{\pi}_{AW}$ متناظر حول $P=\frac{1}{2}$ وليس تباين $\hat{\pi}_{AU}$ كذلك، وتقدّم القيمة الصغيرة لِـ P ، في هذه الطريقة، إجابات قليلة على السؤال الحسّاس).

وإذا كان من الضروري تقدير كل من π_A و π_U فيمكن أن نأخذ عيّنتين عشوائيتين حجماهما n_1 و n_2 بنسب مختلفة من أجل السؤال الحسّاس، P_1 و P_2 ولنرمز بِـ ϕ_1 و ϕ_2 لنسب الأجوبة «نعم» في المجتمعين المعرّفين باختيار القيمتين P_1 و P_2 فعندئذ،

$$\phi_1 = P_1\pi_A + (1-P_1)\pi_U \tag{13.74}$$

$$\phi_2 = P_2\pi_A + (1-P_2)\pi_U \tag{13.75}$$

وتقترح هاتان العلاقتان التقدير،

$$\hat{\pi}_{AU} = \frac{[\hat{\phi}_1(1-P_2) - \hat{\phi}(1-P_1)]}{(P_1 - P_2)} \quad (13.76)$$

حيث،

$$V(\hat{\pi}_{AU}) = \frac{1}{(P_1 - P_2)^2}\left[\frac{\phi_1(1-\phi_1)(1-P_2)^2}{n_1} + \frac{\phi_2(1-\phi_2)(1-P_1)^2}{n_2}\right] \quad (13.77)$$

وإذا كان $P_1 > \frac{1}{2}$ فقد بيّن Greenberg وآخرون (1967) أن هذا التباين يكون أصغر ما يمكن عندما $P_2=0$ أي عندما يُسأل جميع من في العيّنة الثانية (n_2) السؤال الغريب π_U . ويوصي Moors (1971) بهذه الطريقة، إلا إن Greenberg وآخرون (1969) يقترح كقاعدة عمل $P_1+P_2=1$ ، وفي حالة اختيار $P_2=0$ فقد يُضعف هذا تعاون المستجيبين. وعندما يكون $P_1=0.8$ ، مثلاً، فإن 80% من العيّنة 1 و 20% من العيّنة 2 سيُسأل في المتوسط السؤال الحسّاس.

ومع القيم المثلى لِـ n_1 و n_2 من أجل $n=n_1+n_2$ معطى، تبيّن متراجحة كوشي ـ شوارتز أن التباين الأصغري الناتج لِـ $\hat{\pi}_{AU}$ هو،

$$V_{min}(\hat{\pi}_{AU}) = \frac{1}{n(P_1 - P_2)^2}[(1-P_2)\sqrt{\phi_1(1-\phi_1)} + (1-P_1)\sqrt{\phi_2(1-\phi_2)}]^2 \quad (13.78)$$

والنسبة $\frac{n_1}{n_2}$ التي تجعل التباين أصغر ما يمكن هي،

$$\frac{n_1}{n_2} = \frac{(1-P_2)}{(1-P_1)}\sqrt{\frac{\phi_1(1-\phi_1)}{\phi_2(1-\phi_2)}} \quad (13.79)$$

ويستدعي هذا الاختيار تقديرات مسبقة لِـ π_A و π_U ولكن القيم المثلى، إلى حد ما، غير حسّاسة لتغيرات بسيطة. ويعطي Greenberg وآخرون (1969) توصيات حول اختيارات P_2 ، n_1 ، P_1 و n_2 . ولحفظ السرّية من المفيد أخذ π_U مساويًا تقريبًا لِـ π_A .

وقد دُرست أشكال كثيرة لهذه الطريقة (مثلاً، استخدام سؤالين لا صلة لهما ببعضهما)، كما دُرست الانحيازات الناتجة في $\hat{\pi}_{AW}$ و $\hat{\pi}_{AU}$ إذا أجاب كسر من المستجيبين عن السؤال بصورة كاذبة. وقد طبّق Greenberg وآخرون (1971) طريقة

العيّنتين لتقدير المتوسط μ_A لمتغير حساس منفصل أو مستمر، وذلك بطرق مشابهة لتلك التي تقـود إلى المعـادلتـين (13.74) و (13.75) . والسؤال الغـريب يقـدّر المتـوسط μ_U لمتغـير غير حسّـاس. وتتلقى زمر جزئية عشوائية تتضمن n_i فردًا $(i=1,2)$ السؤال الحسّاس باحتمال P_i والسؤال غير الحسّاس باحتمال $(1-P_i)$ ولذلك فإن المتغيّر المسجّل z_i من أجـل فرد يتبـع خليطًا من توزيعـين بنسبتـين P_i و $(1-P_i)$ وأحـد التوزيعات له متوسط μ_A وتباين $\sigma_A{}^2$ وللآخر متوسط μ_U وتباين $\sigma_U{}^2$ وبالتالي،

$$E(z_i) = P_i\mu_A + (1-P_i)\mu_U \tag{13.80}$$

$$V(z_i) = P_i\sigma_A{}^2 + (1-P_i)\sigma_U{}^2 + P_i(1-P_i)(\mu_A - \mu_U)^2 \tag{13.81}$$

وبصورة مماثلة لِـ (13.76) يكون تقدير μ_A ،

$$\hat{\mu}_{AU} = \frac{[(1-P_2)\bar{z}_1 - (1-P_1)\bar{z}_2]}{(P_1 - P_2)} \tag{13.82}$$

ومن أجل فعالية عظمى نحتاج إلى الشروط $\mu_U = \mu_A$ ، $\sigma_U{}^2 = 0$ بينما تكون الشروط $\mu_U = \mu_A$ ، $\sigma_U{}^2 = \sigma_A{}^2$ أفضل من أجل حفظ السرّية.

ويعطي Warner (1971) إطارًا نظريًا لصف واسع من نماذج الإجابة العشوائية. وكما تقترح (13.74) و (13.75) فالحيلة هي تقدير دوال خطية معيّنة في المعالم الحسّاسة والغريبة π أو μ بعدد من المعادلات يساوي عدد المعالم التي يراد تقديرها.

وقـد طُبقت الـطريقـة للحصول على تقديرات لنسب المواليد غير الشرعيين، حالات الإجهـاض، مستخدمي الهيروين، أشخاص على تماس بالجريمة المنظمة، ومتوسط الدخل مع عدد الإجهاضات كتطبيقات على متغير مستمر ـ منفصل. وبما أن الـطريقـة جذبت انتبـاهًا على نطاق واسع، فمن المحتمل ظهور تطبيقات إضافية، ويقدّم Horvitz ، Creenberg و Abernathy (1975) مراجعة ممتازة لهذا الموضوع.

وقـد ظهرت حديثًا مناقشات لدرجة السرّية التي يمتلكها المستجيب في نسخ مختلفة من طريقة الزيارة المعشّاة. وفي بعض النسخ قد يكون المعاين قادرًا على تخمين واقع بعض المستجيبين بالنسبة للمسألة الحسّاسة ويكون التخمين صحيحًا باحتمال عال إلى حد ما وهي إحدى المقوّمات غير المرغوبة لهذه الطريقة.

(١٣-١٩) خلاصة

من حيث تأثيرها على العلاقات المعطاة في الفصول السابقة، يمكن تصنيف الأخطاء التي لا تعود إلى المعاينة كما يلي:

١ - في حالة عدم التغطية وعدم الاستجابة، تكون النتيجة الأكثر أهمية هي إمكانية أن تصبح التقديرات منحازة، وذلك بسبب أن الجزء من المجتمع الذي لم تجر تغطيته قد يختلف عن الجزء الذي جرت معاينته. وتوجد الآن دلالة واسعة على أن هذه الانحيازات تختلف بشكل كبير من مفردة إلى مفردة ومن مسح إلى مسح، وتكون أحياناً كبيرة وأحياناً مهملة. والنتيجة الثانية هي أن التباينات تزداد، بالطبع، لأن العيّنة التي حصلنا عليها فعلاً أصغر من العيّنة الهدف، ويمكن أخذ هذا العامل بعين الاعتبار بصورة تقريبية، على الأقل، عند اختيار حجم العيّنة الهدف.

٢ - أخطاء القياس المستقلة من وحدة إلى وحدة ضمن العيّنة، والتي متوسطها فوق المجتمع يساوي الصفر، أُخذت بعين الاعتبار بشكل مناسب، وذلك في العلاقات المعتادة الخاصة بحساب الأخطاء المعيارية للتقديرات، شريطة أن تكون حدود الـ ت م م مهملة. وتخفّض مثل هذه الأخطاء دقة التقديرات، ومن الجدير أن نبحث عما إذا كان هذا التخفيض جديًا.

٣ - إذا كانت أخطاء القياس في وحدات العيّنة المختلفة مرتبطة فالعلاقات المعتادة للأخطاء المعيارية تكون منحازة. ومن المحتمل أن تكون الأخطاء المعيارية صغيرة جدًّا باعتبار أن معظم الارتباطات هي في التطبيق العملي، إيجابية. وقد يُغفل هذا النوع من الإزعاج بسهولة، وفي الغالب يمكن أن يمرّ دون أن يلحظه أحد.

٤ - انحياز ثابت يؤثر في جميع الوحدات على حدّ سواء وهو الأصعب اكتشافًا من الجميع. وسوف لا تكشف أية معالجات لبيانات العيّنة مثل هذا الانحياز.

وكما أشرنا في هذا الفصل فإن دراسة هذه المسائل بطيئة وصعبة. إلا أن بداية طيبة قد تحققت. وظهر الكثير من البراعة في ابتكار تقنيات لتقويم الأخطاء التي لا تعود إلى المعاينة والتحكم بها. ومع أنه يصعب الوصول إلى تعميمات واسعة في هذا الميدان، إلا أن المعلومات تتراكم حول طبيعة وحجوم أخطاء القياس في أنواع مختلفة من المسوح.

ونتعلّم المـزيد أيضًـا حول ما يمكن إنجـازه من خلال التـدريب الجيد للمعـاينـين والإشراف الجيّد عليهم، ومن خلال الاختيار المسبق، ومـع آليات التحكم بنـوعية العمل الميداني، وتثمين النجاحات ونقاط الضعف في العملية بعد الانتهاء من المسح .

وتحت فروض معيّنة، يُقدم قياس ثان لعيّنة جزئية من الوحدات من قِبل معاين آخر ذي مهارة مماثلة تقديرات لتباين الإجابة البسيط ونسبته إلى تباين الإجابة الكلي، بالإضافة إلى تقدير تقريبي لنسبته إلى تباين المعاينة الذي تخضع له هذه العيّنة الجزئية . ويزوّدنا ابتكار ثان ـ العيّنات الجزئية المتداخلة ـ بتقديرات للتباين الكلي (تباين المعاينة مضـافًـا إليه تبـاين الإجابة)، وبتقدير لمركبة الارتباط في تباين الإجابة . أما تركيب التداخل والقياس المكرّر فهما على وجه الخصوص مثمران .

تمـاريـــن

(١٣-١) في طرق ميدانية مختلفـة الفعالية يمكن جعل طبقة «الإجابة» مؤلفة من 60 ، 80 ، 90 أو 95% من المجتمـع بكامله . ومن أجل النسبة المئوية التي نريد تقـديرهـا، نعلم أن المتـوسـطات الحقيقية لطبقـة الإجـابـة هي : الـطبقـة 60% متـوسـطهـا 40.7 ، الـطبقـة 80% متـوسـطها 43.5 ، الطبقة 90% متوسطها 44.8 ، الطبقة 95% متوسطها 45.4 والـ 5% الأخيرة متوسطها 59.0 (أ) من أجل طريقة تُعاين فقط طبقة الـ 60% بيّن أن الجذر التربيعي لمتوسط مربعات الخطأ للنسبة المئوية المقدّرة هو،

$$\sqrt{(2414/n)+28.94}$$

حيث n عدد الاستبيانـات المستكملة التي حصلنـا عليهـا . (ب) بيّن أنـه لا يمكن الـوصـول إلى جذر متـوسط مربعـات خطأ الـ 5% بطريقـة تستخـدم طبقـة الـ 60% مستجيبًـا، إلا إنـه يمكن الحصـول عليهـا بما يزيد قليلًا على 100 استبيان مستكمل، وذلك من أجل طرق الـ 80% مستجيبًا أو أفضل . (جـ) إذا طُلب جذر متوسط مربعات الخطأ لِـ 2% فما هي الطرق التي يمكن استخدامها لبلوغ ذلك، وما هو حجم العيّنة الذي نحتاجه؟

(١٣-٢) في (١٣-١) (جـ) لنفرض أن تكلفة الاستبيان الكامل هي 5 دولارات من أجل الطريقة الميدانية التي تمتلك 90% مستجيبًا. والحصول على استبيان كامل من الـ 5% التالية من المجتمع يكلّف 20 دولارًا. فمن أجل جذر متوسط مربعات الخطأ لِـ 2% هل يكون من الأوفر استخدام طريقة تؤمن نسبة إجابة 95% ؟

(١٣-٣) يتألف مجتمع من طبقتين متساويتي الحجم. واحتمال العثور على مستجيب في منزله ويرحّب بالمقابلة في أي زيارة هو 0.9 لأفراد الطبقة الأولى و 0.4 لأفراد الطبقة الثانية. (أ) وفقًا لرموز الفقرة (١٣-٥) بيّن أن،

$$w_{i1} = 1-(0.1)^i, \qquad w_{i2} = 1-(0.6)^i$$

(ب) إذا كان حجم العيّنة الأصلي n_o فاحسب العدد الكلي المتوقع للمقابلات التي نحصل عليها من 1 ، 2 ، 3 ، 4 و 5 زيارات. (جـ) إذا كانت التكاليف النسبية لكل مقابلة تمّت في الزيارة الـ i هي 100 ، 120 ، 150 ، 200 و 300 من أجل i مساوية لِـ 1 ، 2 ، 3 ، 4 و 5 على الترتيب. فاحسب متوسط تكلفة المقابلة الواحدة من أجل المقابلات التي تمت حتى الزيارة i . (د) الأموال المتوافرة للمسح كافية لتمويل استكمال 300 من الزيارات الأولى. وإذا كانت سياسة العمل هي الإصرار على i زيارة فما هي الأعداد الكلية المتوقعة للمقابلات المستكملة التي يمكننا الحصول عليها من أجل المبلغ المتوافر نفسه، وذلك عندما يكون $i=1,2,3,4,5$ ؟

(١٣-٤) في التمرين (١٣-٣) إذا كان للأشخاص في الطبقة 1 متوسط 40% من أجل نسبة مئوية، نقوم بتقديرها، لتوزيع ثنائي ما؛ وكان للأشخاص في الطبقة 2 متوسط 60% . (ا) احسب الانحياز في متوسط العيّنة في حالة i تساوي 1 ، 2 ، 3 ، 4 و 5 زيارات. (ب) احسب تباينات متوسطات العيّنة في حالة التكاليف الموصوفة في الجزء (د) من التمرين (١٣-٣). (ولتوفير الجهد الحسابي، يمكن أخذ التباين على أنه $2600/n_i$ ، حيث n_i العدد الكلي المتوقع للزيارات التي نحصل عليها. (جـ) أية سياسة تعطي أصغر MSE ؟

(١٣-٥) في الفقرة (١٣-٦) (عيّنة جزئية من غير المستجيبين) تحقق من العلاقة (صفحة ٥٣٢) الخاصة بنسبة التكلفة المتوقعة للحصول على V محدد بدون معاينة جزئية إلى التكلفة المتوقعة الصغرى،

$$\text{Ratio} = \frac{F(c_0 + c_1 W_1 + c_2 W_2)}{[\sqrt{(F - W_2)(c_0 + c_1 W_1)} + W_2\sqrt{c_2}]^2}$$

حيث $F=S^2/S_2^2$. ليكـــن $c_2=16, c_0=c_1=1$. (ا) إذا كان $F=1$ ، فبـيّن أن نسبة التكلفة تفترض القيمة العظمى 1.25 عندما $W_2=0.2$ أو 0.25 (ب) إذا كان $F=1.5$ فبيّن أن القيمة العظمى هي 1.41 من أجل $W_2=0.3$ أو 0.35 .

(١٣-٦) في مسح يتعلق بالدواجن والخنازير التي يُحتفظ بها في حدائق وبعض الأراضي المستأجرة كزرائب (1957, Gray) جرى استفسار بريدي مع التذكير به عدة مرّات وتبعته مقابـلات لعيّنـة جزئية من غير المستجيبين. ومن خلال حكم مُسبق اختير $k=2$ (أي 50% عيّنة جزئية) . وقد توافرت بعد المسح المعلومات التالية عن إحدى المفردات المهمة، وذلك وفق رموز التمرين (١٣-٥) .

$$\frac{c_1}{c_0} \doteq 0.15, \qquad \frac{c_2}{c_0} \doteq 9.5, \qquad W_1 \doteq 0.8, \qquad S^2 \doteq S_2{}^2$$

بإيجاد VC في حالة $k=2$ وفي حالة القيمة المثلى لِـ k ، حدّد ما إذا كان $k=2$ اختيارًا جيّدًا .

(١٣-٧) في مسح على طريقة Politz-Simmons وُجد 350 مستجيبًا في المنزل في الزيارة الأولى وذلك من عيّنة ابتدائية حجمها 660 . وكان عدد الذين أفادوا بأنهم كانوا في المنـزل في 0,1,...,5 من الأمسيات السابقة وعدد الذين أجابوا بنعم على سؤال في المسح كانت كما يلي :

	0/5	1/5	2/5	3/5	4/5	5/5
العدد	14	35	55	74	94	118
الأجوبة نعم	4	13	20	30	42	156

احسب تقدير Politz-Simmons لنسبة الأجوبة «نعم» في المجتمع وقارنه بالتقدير الثنائي البسيط .

(١٣- ٨) يتضمن مجتمع فيه $N=6$ ثلاث وحدات يكون الجواب الصحيح فيها عن سؤال هو «نعم» وثلاث وحدات يكون الجواب «لا» . وبالنظر إلى أخطاء القياس فإن احتمال الحصول على الإجابة «نعم» من وحدة «نعم» هو 0.9 . واحتمال الحصول

على إجابة «لا» من وحدة «لا» هو أيضًا 0.9 . (ا) باستخدام توزيع كل الإجابات الممكنة في حالة عيّنات حجمها 2 ، بيّن أن احتمالات أن تعطي العيّنة 0 ، 1 و 2 من إجابات الـ«نعم» هي 0.218 ، 0.564 و 0.218 . (ب) بيّن أن تباين النسبة المقدّرة لإجابات الـ«نعم» هو 0.1090 تحقق من النتيجتين (13.40) و (13.41) في الفقرة (١٣-١٠) . (جـ) ماذا يمكن أن يكون تباين النسبة المقدّرة لإجابات الـ«نعم» إذا لم يكن هناك أخطاء قياس؟

(١٣-٩) في جزء من استبيان العمل البنغالي عام 1942 (Mahalanobis, 1946) أُخذت عيّنة عشوائية من حوالي 175 أسرة من كلٍ من ثلاث طبقات. وقُسمت العيّنة من كل طبقة إلى خمس عيّنات جزئية عشوائية، وخُصّصت كل منها إلى معاين مختلف. وقد عمل المعاينون الخمسة في كافة الطبقات الخمس. ومن أجل نفقات الطعام كان الجزء المناسب من تحليل التباين (على أساس أسرة بمفردها) كما يلي:

	df	ms	E(ms)
ما بين المعاينين	4	22.3	$\sigma_{\mu\cdot}^2+\sigma_d^2+35\sigma_{IS}^2+105\sigma_I^2$
المعاينون × الطبقات	8	9.6	$\sigma_{\mu\cdot}^2+\sigma_d^2+35\sigma_{IS}^2$
ما ضمن العيّنات الجزئية	510	9.9	$\sigma_{\mu\cdot}^2+\sigma_d^2$

وإذا كانت w_{hi}, g_i تمثل انحيازات المعاين i ، فالنموذج من أجل أسرة بمفردها هو،

$$y_{hij\alpha}=\bar{\mu}_h+g_i+w_{hi}+(\mu'_{hij}-\bar{\mu}_h)+d_{hij\alpha}$$

التباينات $\sigma_I^2 \quad \sigma_{IS}^2 \quad \sigma_{\mu\cdot}^2 \qquad \sigma_d^2$

تحقق من صحة العبارات المعطاة لِـ E(ms) وقدّر نسبة التباين الكلي للمتوسط التي يمكن أن نعزوها إلى انحيازات العدّاد.

(١٣-١٠) لنأخذ فعلاً غير مشروع ارتكبه 10% من مجتمع (π_A=0.1) . إذا أجاب جميع المستجيبين بصدق، فقارن من أجل n=500 قيم $V(\hat{\pi}_A)$ المعطاة بِـ (ا) سؤال حسّاس مباشر، (ب) طريقة Warner حيث P=0.8 ، (جـ) طريقة السؤال الغريب مع π_u=0.2 معروف، (د) طريقة السؤال الغريب في عيّنتين مع π_u مساوية فعلاً لِـ 0.2 ، إلا إنها مجهولة، وذلك عندما يكون P_1=0.8 ، P_2=0 . كما يوصي

Moors ، (هـ) الطريقة نفسها ولكن عندما $P_1=0.8$ ، $P_2=1-P_1$ افترض أنه يمكنك استخدام القيمة المثلى لِـ n_1/n_2 في (د) و(هـ). ويمكن تفادي بعض الأرقام العشرية بحساب $V(100\hat{\pi}_A)=10^4V(\hat{\pi}_A)$ في الطرق المذكورة.

(١٣-١١) في التمرين (١٣-١٠)، لنفرض أن جميع المستجيبين أجابوا بصدق في أي من طرق الإجابة المعشّاة (ب)، (جـ)، (د)، أو (هـ)، إلا إنه في الطريقة (ا) حيث يُوجّه بصورة مباشرة سؤال حساس واحد، قام بعض المستجيبين بإنكار ارتكابهم للفعل مع إنهم ارتكبوه. في حالة $n=500$ ما هي طرق الإجابة المعشّاة التي تعطي $MSE(\hat{\pi}_A)$ أصغر من طريقة (ا) إذا أنكر ارتكاب الفعل تحت الطريقة (ا) بنسبة (i)15% ، (ii)20% ، (iii)25% ممن ارتكبوه؟

أجوبة التمارين

(١ - ١) (ا) أمثلة عن مشكلات التعريف هي قرارات ما إذا كنا سنعدّ الكلمات في مقدمة أو في فهرس، وكيف سنتعامل مع الرموز الرياضيّة «ككلمات». وعلى أي حال، ففي كتاب كهذا الكتاب يتضمن الكثير من الرموز الرياضيّة، يبدو من غير المحتمل أن يُطلب تعداد كلمات، سواء اشتملت الكلمات على الرموز أم لا.

(ب) (١) تشكّل الصفحات إطارًا مريحًا. ومن عيوب الصفحة (كوحدة معاينة نقوم بعدّ جميع كلماتها إذا كانت في العيّنة)، أنّه مع العديد من الأشكال التوضيحيّة سيكون عدد الكلمات للصفحة الواحدة متغيرًا تغيرًا ملحوظًا بسبب الصفحات غير الكاملة. وقد يكون من الجدير أولًا وضع قائمة بجميع الصفحات غير الكاملة، وتشكيل مجتمعين جزئيين أو طبقتين، واحد للصفحات غير التامّة والآخر للصفحات التامة، ومعاينتها بصورة منفصلة بطرق المعاينة الطبقيّة الموصوفة في الفصل الخامس.

(٢) وإحدى مشكلات السطر هي أنّ الحصول على قائمة بالأسطر بحيث يمكن معاينتها مباشرة يستغرق الكثير من الوقت، بالإضافة إلى وجود سطور غير تامّة في نهاية فقرات. وبما أن عدد الكلمات في السطر ينبغي أن يكون مستقرًا إلى حد ما، فقد يكون الحل، على أي حال، في استخدام المعاينة على مرحلتين (الفصل ١١)، فنسحب أولًا عددًا

من الصفحات ثم نحصي عدد السطور في كل صفحة نختارها، ثم نسحب عيّنة جزئيّة من السطور من كل من هذه الصفحات.

(١ ـ ٢) يفترض هذا السؤال أن نسحب أولاً عيّنة من البطاقات باحتمالات متساوية.

(ا) إذا أهملنا الأسماء في العيّنة التي لا تنتمي إلى المجتمع الهدف، فالمشكلة الوحيدة هي أن حجم العيّنة من الأسماء من المجتمع الهدف ستكون بصورة عامة أقل من عدد البطاقات، وستكون متغيرًا عشوائيًا يعتمد على البطاقات التي يتفق أن تكون قد اختيرت.

(ب) المشكلة هي أن الأسماء التي تظهر على عدة بطاقات سيكون لها احتمالات أعلى في الاختيار. وإحدى طرق تناول هذه المشكلة هي إحصاء عدد البطاقات التي يظهر عليها اسم جرى اختياره، واستخدام هذا الاسم للقيام بالتقدير مستخدمين طرق الاختيار المناسبة باحتمالات غير متساوية (الفصل ٩ ا). والطريقة الأخرى التي تمنح كل اسم فرصًا متساوية ولكنها يمكن أن تنطوي على العديد من الاختيارات المرفوضة هي أن نحتفظ بالبطاقة فقط إذا كانت الأولى من مجموعة البطاقات التي تتضمن الاسم نفسه.

(جـ) كما في (ب) يجري اختيار الأسماء باحتمالات غير متساوية. ولا أعلم طريقة سهلة تمنح كل اسم الفرصة نفسها. وإذا كان عدد البطاقات التي تتضمن الاسم نفسه مسجلة بطريقة ما فيمكن استخدام طريقة الاحتمالات غير المتساوية كما في (ب).

(١ ـ ٣) بعض الاقتراحات

(ا) دليل حديث لمحلات بيع الحقائب ومتاجر المنوّعات.

(ب) أمكنه حفظ المواد المفقودة والتي تشرف عليها شركات النقل الداخلي بالحافلات أو بقطارات الأنفاق.

(جـ) المستشفيات والعيادات الخاصة في المنطقة الجغرافية التي تقع فيها عضة أفعى، بالإضافة إلى أي منظّمة صحيّة عامّة تُشكّل مرجعًا إجباريًا للإفادة عن وقوع عضة أفعى. ونقاط الضعف في الأطر الثلاثة جميعها يمكن أن تتمثل في عدم الكمال أو النقص، بالإضافة إلى الكلفة العالية في (جـ) إذا كانت عضّات الأفاعي نادرة ولا يجري التبليغ عنها إلى جهة مركزية.

(د) تُستخدم غالبًا قائمة من المنازل كإطار لاختيار عيّنة من الأسر. ومع أنه سيوجد بعض النقص (أسر لا يمكن الوصول إليها) فقد تكون المشكلة الرئيسة في أخطاء القياسات.

(١ - ٤) إحدى المشكلات هي النقص (عدم الكمال) بسبب المنشآت الجديدة. وفي عيّنة من العناوين، يمكن للمعاين أن يعالج عادة مشكلة المساكن الجديدة، إذ يتحقق، من أجل كل عنوان في العيّنة مما إذا كان هناك مساكن جديدة بين هذا العنوان والعنوان الذي يليه في الدليل، وفي حال وجود مثل هذه المساكن الجديدة فإنه يضمّها إلى العيّنة. وقد لا يتضمن الدليل مناطق بكاملها من المنشآت الجديدة مما يستدعي تطوير إطار منفصل. وسحب قائمة من العناوين مفضّل على سحب قائمة من الأشخاص، باعتبار أن العناوين أكثر ديمومة. إلا إنّه، ولأسباب تتعلق بنفقات السفر، يمكن أن تكون وحدة المعاينة جادّة في مدينة ويُسحب من الجادّة عيّنة جزئيّة من المساكن.

(١ - ٥) $ 80390 و $ 82970 .

(١ - ٦) احتمال الثقة هو حوالي 0.054 (نحسبه من $t=-1.67$ بخمس وعشرين درجة حرية). ويفترض هذا أن الإيصالات المستقبليّة تتبع توزيع التكرار نفسه الذي تتبعه العيّنة من 26 إيصالاً، وأن هذا التوزيع طبيعي.

(١ - ٧) عندما يعود الـMSE بكامله إلى الانحياز، يكون التقدير مخطئًا دائمًا بمقدار $1\sqrt{MSE}$. وبالتالي يكون احتمال أن يتجاوز الخطأ أو يساوي $1\sqrt{MSE}$ هو الواحد، واحتمال أن يتجاوز الخطأ أو يساوي $1.96\sqrt{MSE}$ ، أو $2.576\sqrt{MSE}$ هو الصفر.

(٢ - ٤) $\hat{Y}= 51,473$ الاحتمال حوالي 0.9

(٢ - ٥) نعم . $\sigma(\hat{Y})$ هو 98.4 .

(٢ - ٦) $\hat{Y}=20,238$ ، $s(\hat{Y})=849$.

(٢ - ٧) (ا) عام : $\hat{R}=15.46$ وخاص : $\hat{R}=12.75$.

(ب) عام : $s(\hat{R})=0.761$. خاص : $s(\hat{R})=0.727$. ومن أجل الـ ت م م نأخذ $f=100/468$.

(جـ) $14.2 < R < 16.7$.

(٢ - ٨) $\dfrac{\text{أثر التصميم (Deff)}}{\text{الانحراف المعياري لأثر التصميم } S.e._{deff}} = \dfrac{2.71}{1.186} = 2.28$.

P يساوي 0.023 تقريبًا. لاحظ عدم استخدام الـ ت م م في حساب الانحراف المعياري لأثر التصميم.

(٢ - ٩) (ا) 9408 ، الانحراف المعياري = 780 .

(ب) 9472 ؛ الانحراف المعياري = 1104 .

(٢ - ١٠) الانحراف المعياري (بالآلاف) يساوي (ا) 800 ، 14 ؛ (ب) 3900 ؛ (جـ) 3140

(٢ - ١١) 9.2 . (ا) 2.7 ؛ (ب) 2.4 .

(٢ - ١٢) (ا) $n=60$ ، بثلاثين من كل ميدان.

(ب)$n=80$ سيؤدي المطلوب إذا تراوح عدد المالكين في العيّنة بين 20 و 60 . ومع $n=80$ فإن احتمال حصول ذلك هو فقط حوالي 0.54 (من جداول التوزيع الثنائي). ومع $n=100$ ، سيكون أي عدد للمالكين ضمن العيّنة بين 19 و 81 كافيًا، واحتمال حدوث هذا هو حوالي 0.94 .

(٢ - ١٤) (ا) 420 ؛ (ب) 940 ؛ (جـ) كلاهما غير منحاز؛ (د) التقدير (ب).

(٣ - ٢) 1066 ، 1334 كما ينتج من التقريب الطبيعي، معادلة (3.19).

(٣ - ٣) حاسم تقريبًا.

(٣ - ٦) (ا) $76.2\pm3.6\%$ ؛ (ب) 1738 ± 280 أسرة.

(٣ - ٧) 1789 ± 268 أسرة.

(٣ - ٨) كنتيجة مضبوطة. $\dfrac{V(\hat{A}_1)}{V(\hat{A}_1^-)} = \dfrac{N_1^2\, n\, Q_1}{N^2 N_1 (1-\pi)(Q_1 + P_1 \pi)}$

والآن $N_1 = N(1-\pi)$ ، وفي عيّنات كبيرة $n_1 \doteq n(1-\pi)$. هذه البدائل تعطي النتيجة المذكورة. ولكي يكون $V(\hat{A}_1)V(\hat{A}_1^-)$ صغيرًا يجب أن يكون $\pi(1-Q_1)/Q_1$ كبيرًا. وهذا يعني أن Q_1 يجب أن يكون صغيرًا: بعبارة أخرى، يجب أن تكون نسبة ما يقع من الميدان 1 في الصف C كبيرة. ذلك لأنه من أجل Q_1 معطاة ينبغي أن تكون π كبيرة.

(٣ - ٩) جميعها تعطي $A_U = 13$. من التوزيع فوق الهندسي، احتمال عدم وجود وحدات من C في العيّنة هو 0.061 من أجل $A_U = 12$ و 0.0434 من أجل $A_U = 13$. ومن التوزيع الثنائي، $P_U = 0.4507$ و $\sqrt{1-f}\,P_U = 0.4114$ ، مما يعطي $A_U = 12.3$. ويعطي المثال ٣ من الصفحة ٨٦-٨٧ القيمة 0.061 من أجل $A_U = 12$ و 0.044 من أجل $A_U = 13$.

(٣ - ١١) يبدو التقدير في (ب) أكثر دقة.

(٣ - ١٢) القيمة الأعلى هي PQ/n بالمقارنة مع PQ/mn من صيغة التوزيع الثنائي. ويقع هذا عندما يتألف كل عنقود بالكامل من المقادير 1 أو بالكامل من المقادير 0 . وأدنى قيمة يمكن أن تكون صفرًا إذا كان كل عنقود يعطي النسبة P نفسها. (وهذا ممكن فقط من أجل قيم معيّنة لـP و m).

(٣ - ١٣) التباين 0.00184 بطريقة النسبة و 0.00160 بتطبيق صيغة التوزيع الثنائي.

(٣ - ١٤) الحجم الوسطي للعيّنة = m/P .

(٤ - ١) 735 منزلاً. نحتاج إلى حجم العيّنة هذا من أجل الأسر التي تمتلك سيارتين إذا كانت $P=10\%$.

(٤ - ٢) حوالي 260 صفحة .

(٤ - ٣) (ا) 2475 ؛ (ب) 4950 .

(٤ - ٤) $n=21$ (آخذين $t=2$) .

(٤ - ٥) $n=484$. من أجل عدد العاطلين عن العمل، يكون معامل الاختلاف حوالي 15% .

(٤ - ٦) 62 عيّنة أخرى .

(٤ - ٧) (ا) $n=278$ ؛ (ب) $n=2315$ ؛ (جـ) $n=3046$.

(٤ - ٨) إذا افترضنا التوزيع المستطيل ضمن كل صف، نأخذ $S^2=0.083h^2$ أو $S=0.29h$. وهذا يعطي التقديرات 230 ، 580 ، 2030 ، 11,600 في الصفوف الأربعة . وإذا استخدمنا التوزيع المثلث القائم في الصف الرابع، نأخذ $S=0.24h$ ، مما يعطي 96000 لهذا الصف .

(٤ - ١١)
$$n_{opt}=\left(\frac{UNS}{2c\sqrt{2\pi}}\right)^{2/3}$$

(٤ - ١٢) (ا) $n=1250$ ؛ (ب) $n=679$. في هذا الجزء، يمكن أن نعطي القيمة 100 + لمتغير دمية y_i من أجل جواب (نعم، لا)، والقيمة 100 - من أجل جواب (لا، نعم)، و 0 فيما عدا ذلك . وعندئذ $\bar{Y}=P_1-P_2$ معبراً عنها كنسبة مئوية . ومع عيّنة مسبقة حجمها $n_1=200$ ، ويمكن استخدام الصيغة (4.7) في الفقرة (٤ - ٧) مما يعطي $n=679$.

(٤-١٣) (ا) احتمال أن يكون لأسرة من أربعة أشخاص 1 ، 2 أو 3 إناث هو تقريبًا $\frac{1}{4}$، $\frac{1}{2}$، و$\frac{1}{4}$، على الترتيب . ولتقدير النسبة P من الإناث تعطي عيّنة عشوائية من n أسرة $V(\hat{P})=0.03125/n$ في مقابل $0.0625/n$ لعيّنة من $4n$ شخصًا . وعامل أثر التصميم deff هو حوالي $\frac{1}{2}$. وفي المثال المقابل في الجدول (٣-٥) مع ثلاثين مسكنًا من أحجام مختلفة، كان عامل أثر التصميم deff المقدّر 0.475 .

(ب) يمكن أن يرتفع عامل أثر التصميم deff بصورة طفيفة من خلال وجود أسر بتوائم متطابقة، باعتبار أن نسب الأسر بإثنين أو بثلاث إناث ستزداد قليلاً .

(٥ - ١) (ا) تعطي المحاصّة النيهانية $n_1=0.87$ ، $n_2=3.13$.

(ب) توجد ثلاثة تقديرات ممكنة تحت المحاصة المثلى وتسعة تحت المحاصّة المتناسبة $V_{opt}(\bar{y}_{st})=\frac{1}{6}=0.167$: $V_{prop}(\bar{y}_{st})=\frac{7}{12}=0.583$

(د) تعطي الصيغة (5.27) $V_{opt}(\bar{y}_{st})=0.159$

(٥ - ٢) (ا) $n_1=375$ ، $n_2=625$.

(ب) $n_1=750$ ، $n_2=250$.

(٥ - ٣) $RP=181\%$ لمحاصّة متناسبة و 214% لمحاصّة مثلى .

(٥ - ٥) عندما نكون $W_1=W_2$ تكون الزيادات النسبية 0.029 من أجل $c_2/c_1=2$ و 0.111 من أجل $c_2/c_1=4$.

(٥ - ٦) (ا) $n_1/n=\frac{1}{3}$ و $n_2/n=\frac{2}{3}$.

(ب) $n=264$ ، $n_1=88$ و $n_2=176$.

(جـ) 1936 $.

(٥ - ٧) (ا) 2288 $ مقابل 1936 $.

(ب) لا . الكلفة الميدانية الدنيا لتخفيض V إلى 1 هي 2230 $.

(٥ - ٨) (ا) $n_1=384$ و $n_2=192$.

(ب) $n_1=400$ و $n_2=1600$.

(جـ) $n_1=1200$ و $n_2=2400$.

(٥ - ٩) الزيادة الكسرية = $\frac{1}{9}$.

(٥ - ١٠) $n_1=541$ ، $n_2=313$ و $n_3=146$.

(٥ - ١٢) في المجتمع 1 ، $V_{prop}=0.143/n$ و $V_{opt}=0.134/n$. وفي المجتمع 2 ، $V_{prop}=0.0491/n$ و $V_{opt}=0.0423/n$. والتخفيض في التباين نتيجة المحاصّة المثلى هو حوالي 6% في المجتمع 1 في مقابل 14% في المجتمع 2 .

(٥ - ١٤) (ا) إذا خمّنا على سبيل التسوية $P_1=45\%$ ، $P_2=25\%$ ، $P_3=7.5\%$ ، وهذا يعطي $n_1=268$ ، $n_2=116$ و $n_3=16$.

(ب) الانحراف المعياري = 0.0225 .

(جـ) الانحراف المعياري = 0.0241 .

(٥ ـ ١٥) عندما يتقارب n إلى N ، نصل إلى مرحلة لا تعود فيها الصيغة القياسيّة $n_h \propto N_h S_h$ لمحاصّة نيمانية مثلى قابلة للتطبيق ؛ باعتبارها ستتطلب $n_h > N_h$ في طبقة واحدة على الأقل. وكما ذكرنا في الفقرة (٥ ـ ٨) لا تعود الصيغة (5.27) صحيحة. يكون الطالب مخطئًا إذا زعم أن (5.27) هي دائمًا خاطئة، وللصيغة مدى محدود إلا أنه يغطي على وجه التقريب جميع التطبيقات.

(٥ ١ ـ ٤) في كل من الحالات الأسوأ نجد $[\Sigma(w_h - W_h)\bar{Y}_h]^2 = (0.105)^2 = 0.0110$. وهكذا فمع الطبقات يكون $MSE(\bar{y}_{st})$ ، كما تعطيه الصيغة (5A.6) مساويًا لـ $0.0108 + 0.0110 = 0.0218$. ومع معاينة عشوائية بسيطة، $V(\bar{y}) = 0.0177$.

(٥ ١ ـ ٦) (ا) $n = 1024$ ، المحاصّة المثلى للمتغير الثاني (المقدار الوسطي المستثمر) تحقق كلي المتطلبين.

(٥ ١ ـ ٧) $W_1 = 0.728$ ، $W_2 = 0.272$ ، $S_1 = 1.806$ ، $S_2 = 4.698$ (في السلّم المرمّز).

(ا) الحجوم المثلى للعيّنات هي $n_1 = 0.507n$ و $n_2 = 0.493n$.

(ب) $V(\bar{y}) = 31.95/n,\ V_{opt}(\bar{y}_{st}) = 6.72/n$

(٥ ١ ـ ٨) (ب) $(b) \int_0^a \sqrt{f(y)}\, dy = \int_0^a \sqrt{2(1-y)}\, dt = 2\sqrt{2}[1-(1-a)^{3/2}]/3$

وبالتالي نريد $[1-(1-a)^{3/2}] = \frac{1}{2}$

(٥ ١ ـ ٩) في التمرين (٥ ١ ـ ٧)، $W_1 = 0.728$ ، كما في قاعدة Dalenius-Hodges ، تكون الأقرب إلى تحقق قاعدة Ekman . وفي المثال (٥ ١ ـ ٨) تعطي قاعدة إكمان $a = (3-\sqrt{5})/2 = 0.38$.

(٥ ١ ـ ١٠) الحل الأمثل هو $L = 7$ من أجل $P = 0.95$ ، $L = 5$ من أجل $P = 0.9$ ، $L = 4$ من أجل $P = 0.8$ و $L = 5$ أو $L = 6$ هو تسوية جيدة.

(٥ ١ ـ ١١)(ا) الكسب في الدقة هو حوالي 110% .

(ب) الكسب من المعاينة الطبقية المتناسبة فوق المعاينة العشوائية البسيطة هو حوالي 90% .

(٥ ١ ـ ١٢)زد n_1 كما يقترح التلميح مبقيًا $n_2 = 400$. نحتاج إلى $n_1 = 140$ ، مما

يعطي $n=540$.

(٦ - ١) من أجل التقدير النسبة $V(\hat{Y}_R) = N^2(1-f)S_d^2/n$ ومن أجل النشر البسيط $V(\hat{Y}) = N^2(1-f)S_y^2/n$ ، حيث $d=(y-Rx)$. من أجل العيّنة من 21 منزلاً نجد تقديرات S_d^2 و S_y^2 كما يلي: عدد الأطفال $s_d^2=0.49$ ، $s_y^2=1.61$ ؛ عدد السيارات، $s_d^2=0.41$ و $s_y^2=0.39$ ؛ عدد أجهزة التلفاز، $s_d^2=0.51$ ، $s_y^2=0.45$. يبدو أن التقدير النسبة متفوق بالنسبة للأطفال.

(٦ - ٢) الكسب 66% . على الأقل 11 وحدة بطريقة النسبة.

(٦ - ٣) الحدود التربيعية (870, 29, 100, 27) ؛ الحدود الطبيعية (700, 29, 030, 27) .

(٦ - ٤) طبق النظرية (٦-٣) على تقدير $R = \bar{Y}/\bar{X}$. وفي حالة عيّنات كبيرة استخدم $\bar{y}/\bar{X}$ إذا كان [ضعف معامل الاختلاف لـ y / معامل الاختلاف لـ x] $\leqslant r$ ، واستخدم $\bar{y}/\bar{x}$ فيما عدا ذلك، حيث r ومعاملات الاختلاف تقديرات عيّنة.

(٦ - ٥) متوسطات مربّعات الخطأ هي 46.5 من أجل التقدير النسبة المنفصل و 40.6 من أجل التقدير النسبة المركب. وفي الحالتين كلتيهما نجد أن مساهمة الانحياز في متوسط مربعات الخطأ يمكن إهمالها.

(٦ - ٦) من أجل طريقة لاهيري، $V(\hat{Y}_{Rt}) = 40.1$

(٦ - ٧) المجموع المقدّر للمجتمع = 116.21 مليونًا. التباين النسبي 0.00111 ، والانحراف المعياري بالتالي هو 3.87 = (116.21) (0.0333) مليونًا. التقدير هو ضمن انحراف معياري واحد من المجموع الحقيقي.

(٦ - ٨) التقديرات هي: (ا) 1896 ، (ب) 1660 ، (جـ) 1689 . في (جـ) نجد $w_1=2.38$ ، $w_2=-1.38$. الانحرافات المعيارية المقدّرة هي: (ا) 256 ، (ب) 36.9 ، (جـ) 18.6 . ومن أجل الانحراف المعياري في (ب) استخدمت الصيغة $\hat{Y}_R\sqrt{(1-f)(c_{yy}+c_{11}-2c_{y1})/n}$ ، حيث $\hat{Y}_R$ هو التقدير النسبة لـ Y ، أي 1660 . ومن أجل الانحراف المعياري في (جـ) استخدمت $\hat{Y}_{MR} = 1689$.

(٧ - ١) التقدير يساوي 11,080 ؛ الانحراف المعياري = 152 (بما في ذلك ت م م).

(٧ - ٢) لا، باعتبار b قريب جدًا من 1 .

(٧ - ٣) $\hat{Y}_{lr} = 28177 \pm 570$. الدقة النسبية 113% .

(٧ - ٤) $27,751 \pm 694$.

(٧ - ٦) من أجل تقدير الفرق، $V(\bar{y}) = S_e^2/n$ ، ومن أجل تقدير الانحدار الخطي $V(\bar{y}_{lr}) = S_e^2 S_v^2 / n(S_e^2 + S_v^2)$. لتقدير الانحدار التباين الأصغر، إلا أن تفوقه ليس ذا بال إذا كان S_e^2/S_v^2 صغيراً.

(٧ - ٧) $MSE(\hat{Y}_{lrs}) = 34.5, Bias^2 = 9.7; MSE(\hat{Y}_{lrc}) = 11.9, Bias^2 = 1.2$

(٧ - ٩) $MSE(\hat{Y}_{Rs}) = 8.9, MSE(\hat{Y}_{Rc}) = 6.7$

(٨ - ١) التبـاينـات هي 8.19 (نَمْطي)، 11.27 (عشوائي بسيط)، 8.25 (طبقية، 2)، 7.46 (طبقية، 1) .

(٨ - ٢) $V_{sys} = 0.00141, V_{ran} = 0.00340$

(٨ - ٣) ينبغي أن تكـون العيّنـة النمـطية متفوقة من أجل نسبة الناس من أصل بولنـدي، باعتبـار أن هذا المتغـير يُظهـر تقسيمًا إلى طبقـات على أساس جغرافي. ومن المتوقع أن لايكون متفوّقًا من أجل نسبة الأطفال لأن فترة المعاينة، 1 في الـ5 ، تتطابق مع الحجم الوسطي للأسرة. والشيء نفسه صحيح، إلى مدى أقل، من أجل نسبة الذكور.

(٨ - ٤) الـتـبـايـنـات كما يلي: ذكـور، $V_{srs}=0.0204$ ، $V_{sys}=0.0216$ ؛ أطـفـال، $V_{srs}=0.0204$ ، $V_{sys}=0.0776$ والمـهـنـيون، $V_{srs}=0.0192$ ، $V_{sys}=0.0016$.

(٨ - ٥) التبـاين الفعلي = 8.19 . الطريقة (ا) تعطي 11.29 . ومن أجل الطريقة (ب) يكون التباين المقدّر من عيّنة بمفردها هو $(1-f)(\bar{y}_{i1} - \bar{y}_{i2})^2/4$ ، حيث $\bar{y}_{i1}$ ، $\bar{y}_{i2}$ هما متوسطا النصفين. والمتوسط هو 3.24 . ومن غير المتوقع أن هناك تقدير جدّي بالنقصان.

(٨ - ٧) كلا التباينين $(k^2-1)/6$.

(٨ - ٨) المعاينة العشوائية البسيطة أفضل ما لم يكن $n=1$ أو $K=1$.

(٨ - ٩) الوحدة الـk من كل K من الوحدات، $V(\hat{\hat{Y}})=362.2$ ؛ Yates ، $MSE(\hat{\hat{Y}})=7.3$ ؛ Sethi ، $V(\hat{\hat{Y}})=21.0$ ؛ Singh وآخرون، $V(\hat{\hat{Y}})=81.0$. والمقدران الأخيران في هذا المثال غير منحازين.

(٩ - ١) التكاليف النسبية لاستخدام الأنواع الأربعة من الوحدات هي 100 ، 90.1 ، 79.7 و 77.8 (آخذين الوحدة من النوع الأول كوحدة قياسية).

(٩ - ٢) الدقة النسبية للأسرة هي 211% من أجل نسبة الجنس و 38% من أجل نسبة من زاروا طبيبًا.

(٩ - ٣) الدقة النسبية للوحدة الكبيرة 0.566 في حالة معاينة عشوائية بسيطة و 0.625 في حالة معاينة عشوائية طبقية.

(٩ - ٥) $(a)\ M=5;\ (b)\ M=1$

(٩ - ٦) ينبغي أن تتناقص M المثلى لأن كلفة السفر، وهي تتغير كتغير $\sqrt{n}$ ، تصبح نسبيًا أقل أهمية مع ازدياد n .

(٩ أ - ١) (a) 34,242; (b) 5534; (c) 6493

(٩ أ - ٣) (ا) إذا كان الانحراف المعياري بين الوحدات الكبيرة في الصف h متناسبًا مع M_h .

(ب) إذا كان الاحتمال متناسبًا مع $\sqrt{M_h}$.

(٩ أ - ٤) (ب) $V(\hat{Y}_{HT})=1.75$ ، $V(\hat{Y}_M)=0.27$ ، $V(\hat{Y}_{RHC})=0.33$ تعتبر y_i/z_i ، في هذه المسألة قليلاً، ولكن أداء $\hat{Y}_{HT}$ غير مرض نسبيًا لأن $\pi_i \neq 2z_i$ في طريقة اختيار العيّنة، ويُنصح بمقدّر مورثي في هذه الطريقة لاختيار العيّنة حيث $V(\hat{Y}_M)=0$ إذا كانت y_i/z_i ثابتة، وذلك كما يتضح من (9A-60) .

(جـ) $V(\hat{Y}_M)/V(\hat{Y}_{ppz})=0.54$ ، بينما $(N-n)/(N-1)=\frac{1}{2}$.

(٩ أ - ٥) $MSE(\hat{Y}^{*}_{HT(A)})=7.06,\ (Bias)^2/MSE=0.065,\ V(\hat{Y}_{SRS})=6.5$

(١٠ - ١) (a) 2.00; (b) 2.13

(١٠ - ٣) $(a)\ 165/n;\ (b)\ 148.5/n;\ (c)\ 132/n$

(١٠ - ٤) (ا) $n=660$ حقلاً؛ $n=530$ حقلاً. يتطلب البروتين حقولاً أقل مما يتطلبه الإنتاج.

(١٠ - ٥) $c_1/c_2=8$

(١٠ - ٧) (*a*) 0.93%; (*b*) 0.51%; (*c*) 0.36%

(١٠ - ٨) (ا) من المناسب أن يكون $m_0=7$ أو $m_0=8$.

(ب) 89% من أجل $m_0=7$ و 93% من أجل $m_0=8$.

(جـ) 86% من أجل $m_0=7$ و 89% من أجل $m_0=8$.

(١١ - ٢) تنخفض الدقة النسبية لـIII إلى II من 3.02 إلى 2.75 . إذا كان اختلاف خطتي معاينة هو بصورة رئيسة في مساهمة ما بين وحداتها في التباين، فإن الدقة النسبية للخطة المتفوقة ستتناقص بصورة عامة كلما ازدادت نسبة تباين ما ضمن الوحدات إلى التباين الكلي.

(١١ - ٣) بكلام تقريبي يمكن القول إن تفسير ذلك هو أن Y_i/z_i أكثر استقرارًا من Y_i/M_i في هذه البيانات. إذا أخذنا $z_i=\frac{1}{33}$ ، $\frac{8}{33}$ ، $\frac{24}{33}$ فإن مساهمة ما بين الوحدات وفقًا للطريقة IV ستنعدم.

(١١ - ٤) التباين الكلي: 0.00504 (I*a*), 0.02358 (II), 0.00554 (III)

(١١ - ٦) النسبة المئوية المقدّرة 2.16± 14.2 . العدد المقدّر 3450±540 .

(١١ - ٧) النسبة المئوية المقدّرة 13.9±2.49 .

(١١ - ٩) العدد الكلي للغرف 29400 ، العدد الكلي للأشخاص 50500 ، الأشخاص للغرفة الواحدة، 1.72 ؛ الانحرافات المعيارية: للعدد الكلي للأشخاص 2440 وللأشخاص للغرفة الواحدة 0.066 .

(١٢ - ١) $n=267$ ، $n'=1320$ أو $n=268$ ، $n'=1280$. $V(p_{st})$ بمحاصّة مثلى هو 6.67 عندما يكون p_{st} بالنسب المئوية. وبمعاينة بمفردها، $V(p)=8.33$.

(١٢ - ٢) $c_n/c_n'>9$

(١٢ - ٣) $n/n'=1/19$

(١٢ - ٤) $n'>16n$

(١٢ - ٥) من الصيغة (12.67) ومتجاهلين $\frac{1}{N}$ نجد الانحراف المعياري = 1.25 .

(١٢ - ٦) المكاسب كنسب مئوية من المناسبة الثانية إلى المناسبة السادسة هي 50 ، 75 ، 91 ، 100 و 105 على الترتيب.

(١٢ ـ ٨) قيم $nV(\bar{y}_2'')/S^2$ و $nV(\bar{y}_2')/S^2$ هي كما يلي : $\mu = \frac{1}{4}, \rho = 0.8: 0.885, 0.875;$ $\mu = \frac{1}{4}\ \rho = 0.9: 0.843, 0.840; \mu = \frac{1}{2}, \rho = 0.8: 0.824, 0.810; \mu = \frac{1}{2}, \rho = 0.9: 0.752, 0.746$.

(١٢ ـ ١٢)(ا) ليكن $y_i=1$ من أجل أي وحدة من الصيغة الأولى و $y_i=0$ فيما عدا ذلك، ولتكن الطبقة الأولى هي الطبقة المؤلفة من الوحدات التي تنتمي إلى الصفة الثانية. ووفقـا لرموز النظريتين 12.1 و 12.2 مع إهمال $\frac{1}{N}$ ، $S^2 = P_1Q_1$ و $S_1{}^2 = P_1P_2/(P_1 + P_2)^2$. النتائج في (ا) تأتي مباشرة من النظريات.

(ب) إذا كانت الكلفـة المتـوقعـة C^* فإن المعـاينـة المضـاعفـة مع $\nu_1 = \frac{1}{2}$ ، $k = 2$ (القيمة المثلى)، تعطي $V(\hat{Y}_{st}) = N^2(0.844)/C^*$ ، بينما تعطي عيّنة عشوائية بسيطة $V(\hat{Y}) = N^2(1.875)/C^*$ ، وهي أكبر بمرتين ونيّف.

(١٣ ـ ١) نسبـة استجـابة 90% مع 1047 استبيانًا مستكملاً أو 95% نسبة استجابة مع (701) استبيانًا مستكملاً.

(١٣ ـ ٢) تكلف الطريقـة بنسبـة استجـابة 90% مبلغ 5235 $. وبنسبة استجابة 95% تكلف 5.7895 $ لكل استبيان مستكمل، أو كلفة إجمالية 4058 $.

(١٣ ـ ٣) (*b*) $0.65n_0, 0.815n_0, 0.8915n_0, 0.9351n_0, 0.9611n_0$; (*c*) 100, 104, 108, 112, 117; (*d*) 300,288, 277, 267, 256.

(١٣ ـ ٤) (ا) الانحياز (كنسبـة مئـوية ٪) = -3.85, -2.15, -1.21, -0.69, -0.40 ؛

(ب) التباينات هي 8.67, 9.03, 9.39, 9.74, 10.16 ؛

(جـ) أربع مكالمات.

(١٣ ـ ٦) نعم. VC من أجل $K=2$ هو حوالي 2% فقط فوق القيمة الدنيا لِـ VC .

(١٣ ـ ٧) تقدير بوليتز ـ سيمونز، 39.7% ؛ الثنائي، 42.3% .

(١٣ ـ ٨) (جـ) التباين = 0.1 .

(١٣ ـ ٩) إذا كان خطأ القياس لكل عدّاد مستقلاً من أسرة إلى أسرة فسيكون تباين متوسط العيّنة $(\sigma_{u}{}^2 + \sigma_d{}^2 + \sigma_{1u}{}^2 + \sigma_1{}^2)/525$ بدلاً من $(\sigma_{u}{}^2 + \sigma_d{}^2 + 35\sigma_{1u}{}^2 + 105\sigma_1{}^2)/525$ تسهم انحيازات العدّادين بحوالي 55% من التباين الإجمالي.

(١٣ ـ ١٠) $10^4 V(\hat{\pi}_A) = V(100\hat{\pi}_A)$ يساوي (ا) 1.80 (مباشر)؛ (ب) 10.69 (وارنر)؛ (جـ) 3.30 (π_u معروف)؛ (د) 5.12 (مُورز)؛ (هـ) 6.30 ($P_2=1-P_1$) .

(١٣ - ١١)(أ) $10^4 MSE(\hat{\pi}_A) = 3.81$ ؛ $\hat{\pi}_{AU}$ متفوق إذا كان π_{L} معروفًا.

(ب) $10^4 MSE(\hat{\pi}_A) = 5.47$ ؛ $\hat{\pi}_{AU}$ بسؤالين هو أيضًا متفوق إذا كان $P_2 = 0$.

(جـ) $10^4 MSE(\hat{\pi}_A) = 7.64$. جميع الطرق متفوقة باستثناء طريقة وارنر الأصلية.

المراجع

References

Armitage, P. (1947). A comparison of stratified with unrestricted random sampling from a finite population. *Biometrika*, **34**, 273–280.

Arvesen, J. N. (1969). Jackknifing U-Statistics. *Ann. Math. Stat.*, **40**, 2076–2100.

Avadhani, M. S. and Sukhatme, B. V. (1973). Controlled sampling with equal probabilities and without replacement. *Int. Stat. Rev.*, **41**, 175–183.

Bailar, B. A. (1975). The effects of rotation group bias on estimates from panel surveys. *Jour. Amer. Stat. Assoc.*, **70**, 23–30.

Bailar, B. A., and Dalenius, T. (1969). Estimating response variance components of the U.S. Bureau of the Census Survey Model. *Sankhya*, **B31**, 341–360.

Barr, A. (1957). Differences between experienced interviewers. *App. Stat.*, **6**, 180–188.

Bartholomew, D. J. (1961). A method of allowing for "not-at-home" bias in sample surveys. *App. Stat.*, **10**, 52–59.

Bartlett, M. S. (1949). Fitting a straight line when both variables are subject to error. *Biometrics*, **5**, 207–212.

Basu, D. (1958). On sampling with and without replacement. *Sankhya*, **20**, 287–294.

Bayless, D. L. (1968). Variance estimation in sampling from finite populations. Ph.D. Thesis, Texas A & M University.

Bayless, D. L., and Rao, J. N. K. (1970). An empirical study of stabilities of estimators and variance estimators in unequal probability sampling ($n = 3$ or 4). *Jour. Amer. Stat. Assoc.*, **65**, 1645–1667.

Beale, E. M. L. (1962). Some uses of computers in operational research. *Industrielle Organisation*, **31**, 51–2.

Bean, J. A. (1970). Estimation and sampling variance in the health interview survey. National Center for Health Statistics, Washington, D.C., Series 2, 38.

Bean, J. A. (1975). Distribution and properties of variance estimators for complex multistage probability samples. National Center for Health Statistics, Washington, D.C., Series 2, 65.

Beardwood, J., Halton, J. H., and Hammersley, J. M. (1959). The shortest path through many points, *Proc. Cambridge Phil. Soc.*, **55**, 299–327.

Bellhouse, D. R. and Rao, J. N. K. (1975). Systematic sampling in the presence of a trend. *Biometrika*, **62**, 694–697.

Belloc, N. B. (1954). Validation of morbidity survey data by comparison with hospital records. *Jour. Amer. Stat. Assoc.*, **49**, 832–846.

Birnbaum, Z. W., and Sirken, M. G. (1950a). Bias due to nonavailability in sampling surveys. *Jour. Amer. Stat. Assoc.*, **45**, 98–111.

Birnbaum, Z. W., and Sirken, M. G. (1950b). On the total error due to noninterview and to random sampling. *Int. Jour. Opinion and Attitude Res.*, **4**, 179–191.

Blythe, R. H. (1945). The economics of sample size applied to the scaling of saw-logs. *Biom. Bull.*, **1**, 67–70.

Booth, G., and Sedransk, J. (1969). Planning some two-factor comparative surveys. *Jour. Amer. Stat. Assoc.*, **64**, 560–573.

Bose, Chameli (1943). Note on the sampling error in the method of double sampling. *Sankhya*, **6**, 330.

Brewer, K. W. R. (1963a). A model of systematic sampling with unequal probabilities. *Australian Jour. Stat.*, **5**, 5–13.

Brewer, K. W. R. (1963b). Ratio estimation in finite populations: Some results deducible from the assumption of an underlying stochastic process. *Australian Jour. Stat.*, **5**, 93–105.

Brewer, K. W. R., and Hanif, M. (1969). Sampling without replacement and probability of inclusion proportional to size. I Methods using the Horvitz and Thompson estimator. II Methods using special estimators. Unpublished manuscript.

Brewer, K. W. R., and Hanif, M. (1970). Durbin's new multistage variance estimator. *Jour. Roy. Stat. Soc.*, **B32**, 302–311.

Brooks, S. (1955). The estimation of an optimum subsampling number. *Jour. Amer. Stat. Assoc.*, **50**, 398–415.

Bryant, E. C., Hartley, H. O., and Jessen, R. J. (1960). Design and estimation in two-way stratification. *Jour. Amer. Stat. Assoc.*, **55**, 105–124.

Buckland, W. R. (1951). A review of the literature of systematic sampling. *Jour. Roy. Stat. Soc.*, **B13**, 208–215.

Burstein, H. (1975). Finite population correction for binomial confidence limits. *Jour. Amer. Stat. Assoc.*, **70**, 67–69.

Cameron, J. M. (1951). Use of variance components in preparing schedules for the sampling of baled wool. *Biomerics*, **7**, 83–96.

Chatterjee, S. (1966). A programming algorithm and its statistical applications. O.N.R. Tech. Rept. 1, Department of Statistics, Harvard University, Cambridge.

Chatterjee, S. (1967). A note on optimum stratification. *Skand. Akt.*, **50**, 40–44.

Chatterjee, S. (1968). Multivariate stratified surveys. *Jour. Amer. Stat. Assoc.*, **63**, 530–534.

Chatterjee, S. (1972). A study of optimum allocation in multivariate stratified surveys. *Skand. Akt.*, **55**, 73–80.

Chung, J. H., and DeLury, D. B. (1950). *Confidence Limits for the Hypergeometric Distribution*. University of Toronto Press, Toronto, Canada.

Cochran, W. G. (1942). Sampling theory when the sampling units are of unequal sizes. *Jour. Amer. Stat. Assoc.*, **37**, 199–212.

Cochran, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann. Math. Stat.*, **17**, 164–177.

Cochran, W. G. (1961). Comparison of methods for determining stratum boundaries. *Bull. Int. Stat. Inst.*, **38**, 2, 345–358.

Cochran, W. G., Mosteller, F., and Tukey, J. W. (1954). *Statistical Problems of the Kinsey Report. American Statistical Association, Washington, D.C., p.* 280.

Coleman, J. S. (1966). *Equality of Educational Opportunity*, U.S. Government Printing Office, Washington, D.C.

Cornell, F. G. (1947). A stratified random sample of a small finite population. *Jour. Amer. Stat. Assoc.*, **42**, 523–532.

Cornfield, J. (1944). On samples from finite populations. *Jour. Amer. Stat. Assoc.*, **39**, 236–239.

Cornfield, J. (1951). The determination of sample size. *Amer. Jour. Pub. Health*, **41**, 654–661.

Cox, D. R. (1952). Estimation by double sampling. *Biometrika*, **39**, 217–227.

Dalenius, T. (1957). *Sampling in Sweden*. Contributions to the methods and theories of sample survey practice. Almqvist and Wicksell, Stockholm.

Dalenius, T., and Gurney, M. (1951). The problem of optimum stratification. II. *Skand. Akt.*, **34**, 133–148.

Dalenius, T., and Hodges, J. L., Jr. (1959). Minimum variance stratification. *Jour. Amer. Stat. Assoc.*, **54**, 88–101.

Das, A. C. (1950). Two-dimensional systematic sampling and the associated stratified and random sampling. *Sankhya*, **10**, 95–108.

David, F. N., and Neyman, J. (1938). Extension of the Markoff theorem of least squares. *Stat. Res. Mem.*, **2**, 105.

David, I. P., and Sukhatme, B. V. (1974). On the bias and mean square error of the ratio estimator. *Jour. Amer. Stat. Assoc.*, **69**, 464–466.

Deming, W. E. (1953). On a probability mechanism to attain an economic balance between the resultant error of non-response and the bias of non-response. *Jour. Amer. Stat. Assoc.*, **48**, 743–772.

Deming, W. E. (1956). On simplifications of sampling design through replication with equal probabilities and without stages. *Jour. Amer. Stat. Assoc.*, **51**, 24–53.

Deming, W. E. (1960). *Sample Design in Business Research*. John Wiley and Sons, New York.

Deming, W. E., and Simmons, W. R. (1946). On the design of a sample for dealer inventories. *Jour. Amer. Stat. Assoc.*, **41**, 16–33.

Des Raj (1954). On sampling with probabilities proportional to size. *Ganita*, **5**, 175–182.

Des Raj (1956a). Some estimators in sampling with varying probabilities without replacement. *Jour. Amer. Stat. Assoc.*, **51**, 269–284.

Des Raj (1956b). A note on the determination of optimum probabilities in sampling without replacement. *Sankhya*, **17**, 197–200.

Des Raj (1958). On the relative accuracy of some sampling techniques. *Jour. Amer. Stat. Assoc.*, **53**, 98–101.

Des Raj (1964). The use of systematic sampling with probability proportional to size in a large-scale survey. *Jour. Amer. Stat. Assoc.*, **59**, 251–255.

Des Raj (1966). Some remarks on a simple procedure of sampling without replacement. *Jour. Amer. Stat. Assoc.*, **61**, 391–396.

Des Raj, and Khamis, S. H. (1958). Some remarks on sampling with replacement. *Ann. Math. Stat.*, **29**, 550–557.

Dowling, T. A., and Shachtman, R. H. (1975). On the relative efficiency of randomized response models. *Jour. Amer. Stat. Assoc.*, **70**, 84–87.

Durbin, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. *Jour. Roy. Stat. Soc.*, **B15**, 262–269.

Durbin, J. (1954). Non-response and call-backs in surveys. *Bull. Int. Stat. Inst.*, **34**, 72–86.

Durbin, J. (1958). Sampling theory for estimates based on fewer individuals than the number selected. *Bull. Int. Stat. Inst.*, **36**, 3, 113–119.

Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika*, **46**, 477–480.

Durbin, J. (1967). Design of multi-stage surveys for the estimation of sampling errors *App. Stat.*, **16**, 152–164.

Durbin, J., and Stuart, A. (1954). Callbacks and clustering in sample surveys: an experimental study. *Jour. Roy. Stat. Soc.*, **A117**, 387–428.

Eckler, A. R. (1955). Rotation sampling. *Ann. Math. Stat.*, **26**, 664–685.

Ekman, G. (1959). An approximation useful in univariate stratification. *Ann. Math. Stat.*, **30**, 219–229.

Erdös, P., and Rényi, A. (1959). On the central limit theorem for samples from a finite population. *Pub. Math. Inst. Hungarian Acad. Sci.*, **4**, 49–57.

Ericson, W. A. (1969). Subjective Bayesian models in sampling finite populations. *Jour. Roy. Stat. Soc.*, **B31**, 195–233.

Evans, W. D. (1951). On stratification and optimum allocations. *Jour. Amer. Stat. Assoc.*, **46**, 95–104.

Fellegi, I. (1963). Sampling with varying probabilities without replacement: rotating and non-rotating samples. *Jour. Amer. Stat. Assoc.*, **58,** 183–201.

Fellegi, I. (1964). Response variance and its estimation. *Jour. Amer. Stat. Assoc.*, **59,** 1016–1041.

Feller, W. (1957). *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, New York, second edition.

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika*, **24,** 428–440.

Finkner, A. L. (1950). Methods of sampling for estimating commercial peach production in North Carolina. *North Carolina Agr. Exp. Stat. Tech. Bull.*, **91**.

Finkner, A. L., Morgan, J. J., and Monroe, R. J. (1943). Methods of estimating farm employment from sample data in North Carolina. *N. C. Agr. Exp. Sta. Tech. Bull.*, **75**.

Finney, D. J. (1948). Random and systematic sampling in timber surveys. *Forestry*, **22,** 1–36.

Finney, D. J. (1949). On a method of estimating frequencies. *Biometrika*, **36,** 233–234.

Finney, D. J. (1950). An example of periodic variation in forest sampling. *Forestry*, **23,** 96–111.

Fisher, R. A. (1958). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh, thirteenth edition, section 21, fourth ed. (1932).

Fisher, R. A., and Mackenzie, W. A. (1922). The correlation of weekly rainfall. *Quart. Jour. Roy. Met. Soc.*, **48,** 234–245.

Fisher, R. A., and Yates, F. (1957). *Statistical Tables for Biological, Agricultural and Medical Research.* Oliver and Boyd, Edinburgh, fifth edition.

Foreman, E. K., and Brewer, K. W. R. (1971). The efficient use of supplementary information in standard sampling procedures. *Jour. Roy. Stat. Soc.* **B33,** 391–400.

Frankel, M. R. (1971). Inference from survey samples. Institute for Social Research, Ann Arbor, Mich.

Fuller, W. A. (1966). Estimation employing post strata. *Jour. Amer. Stat. Assoc.*, **61,** 1172–1183.

Fuller, W. A. (1970). Sampling with random stratum boundaries. *Jour. Roy. Stat. Soc.*, **B32,** 209–226.

Fuller, W. A., and Burmeister, L. F. (1972). Estimators for samples selected from two overlapping frames. *Proc. Soc. Stat. Sect. Amer. Stat. Assoc.*, 245–249.

Gallup, G. (1972). Opinion polling in a democracy. *Statistics, a Guide to the Unknown.* J. M. Tanur et al. (eds.), Holden-Day, Inc., San Francisco, 146–152.

Gautschi, W. (1957). Some remarks on systematic sampling. *Ann. Math. Stat.*, **28,** 385–394.

Godambe, V. P. (1955). A unified theory of sampling from finite populations. *Jour. Roy. Stat. Soc.*, **B17,** 269–278.

Goodman, L. A., and Hartley, H. O. (1958). The precision of unbiased ratio-type estimators. *Jour. Amer. Stat. Assoc.*, **53,** 491–508.

Goodman, R., and Kish, L. (1950). Controlled selection—a technique in probability sampling. *Jour. Amer. Stat. Assoc.*, **45,** 350–372.

Graham, J. E. (1973). Composite estimation in two cycle rotation sampling designs. *Comm. in Stat.*, **1,** 419–431.

Gray, P. G. (1955). The memory factor in social surveys. *Jour. Amer. Stat. Assoc.*, **50,** 344–363.

Gray, P. G. (1957). A sample survey with both a postal and an interview stage. *App. Stat.*, **6,** 139–153.

Gray, P. G., and Corlett, T. (1950). Sampling for the social survey. *Jour. Roy. Stat. Soc.*, **A113,** 150–206.

Greenberg, B. G., et al. (1969). The unrelated question randomized response model: Theoretical framework. *Jour. Amer. Stat. Assoc.*, **64,** 520–539.

Greenberg, B. G. et al. (1971). Application of the randomized response technique in obtaining quantitative data. *Jour. Amer. Stat. Assoc.*, **66,** 243–250.

Grundy, P. M., Healy, M. J. R., and Rees, D. H. (1954). Decision between two alternatives—how many experiments? *Biometrics*, **10,** 317–323.

Grundy, P. M., Healy, M. J. R., and Rees, D. H. (1956). Economic choice of the amount of experimentation. *Jour. Roy. Stat. Soc.*, **B18,** 32–55.

Hagood, M. J., and Bernert, E. H. (1945). Component indexes as a basis for stratification. *Jour. Amer. Stat. Assoc.*, **40,** 330–341.

Hájek, J. (1958). Some contributions to the theory of probability sampling. *Bull. Int. Stat. Inst.*, **36,** 3, 127–134.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Pub. Math. Inst. Hungarian Acad. Sci.*, **5,** 361–374.

Haldane, J. B. S. (1945). On a method of estimating frequencies. *Biometrika*, **33,** 222–225.

Hansen, M. H., et al. (1951). Response errors in surveys. *Jour. Amer. Stat. Assoc.*, **46,** 147–190.

Hansen, M. H., and Hurwitz, W. N. (1942). Relative efficiencies of various sampling units in population inquiries. *Jour. Amer. Stat. Assoc.*, **37,** 89–94.

Hansen, M. H., and Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *Ann. Math. Stat.*, **14,** 333–362.

Hansen, M. H., and Hurwitz, W. N. (1946). The problem of nonresponse in sample surveys. *Jour. Amer. Stat. Assoc.*, **41,** 517–529.

Hansen, M. H., and Hurwitz, W. N. (1949). On the determination of the optimum probabilities in sampling. *Ann. Math. Stat.*, **20,** 426–432.

Hansen, M. H., Hurwitz, W. N., and Bershad, M. (1961). Measurement errors in censuses and surveys. *Bull. Int. Stat. Inst.*, **38,** 2, 359–374.

Hansen, M. H., Hurwitz, W. N., and Gurney, M. (1946). Problems and methods of the sample survey of business. *Jour. Amer. Stat. Assoc.*, **41,** 173–189.

Hansen, M. H., Hurwitz, W. N., and Jabine, T. B. (1963). The use of imperfect lists for probability sampling at the U.S. Bureau of the Census. *Bull. Int. Stat. Inst.*, **40,** 1, 497–517.

Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953). *Sample Survey Methods and Theory*. John Wiley and Sons, New York, Vols. I and II.

Hansen, M. H., Hurwitz, W. N., Nisselson, H., and Steinberg, J. (1955). The redesign of the census current population survey. *Jour. Amer. Stat. Assoc.*, **50,** 701–719.

Hansen, M. H., Hurwitz, W. N., and Pritzker, L. (1965). The estimation and interpretation of gross differences and the simple response variance. *Contributions to statistics presented to Professor P. C. Mahalanobis*. Pergamon Press, Oxford, and Statistical Publishing Society, Calcutta, 111–136.

Hansen, M. H., and Waksberg, J. (1970). Research on non-sampling errors in censuses and surveys. *Rev. Int. Stat. Inst.*, **38,** 318–332.

Hanson, R. H., and Marks, E. S. (1958). Influence of the interviewer on the accuracy of survey results. *Jour. Amer. Stat. Assoc.*, **53,** 635–655.

Hartley, H. O. (1946). Discussion of paper by F. Yates. *Jour. Roy. Stat. Soc.*, **109,** 37.

Hartley, H. O. (1959). *Analytic Studies of Survey Data*. Istituto di Statistica, Rome, volume in honor of Corrado Gini.

Hartley, H. O. (1962). Multiple frame surveys. *Proc. Soc. Stat. Sect. Amer. Stat. Assoc.*, 203–206.

Hartley, H. O. (1974). Multiple frame methodology and selected applications. *Sankhya*, **C36,** 99–118.

Hartley, H. O., and Hocking, R. (1963). Convexing programming by tangential approximation. *Management Science*, **9,** 600–612.

Hartley, H. O., and Rao, J. N. K. (1962). Sampling with unequal probabilities and without replacement. *Ann. Math. Stat.*, **33,** 350–374.

Hartley, H. O., and Rao, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika*, **55,** 547–557.

Hartley, H. O., and Rao, J. N. K. (1969). A new estimation theory for sample surveys, II. In *New Developments in Survey Sampling*, N. L. Johnson and H. Smith (eds.), Wiley-Interscience, New York, 147–169.

Hartley, H. O., Rao, J. N. K., and Kiefer, G. (1969). Variance estimation with one unit per stratum. *Jour. Amer. Stat. Assoc.*, **64,** 841–851.

Hartley, H. O., and Ross, A. (1954). Unbiased ratio estimates. *Nature*, **174,** 270–271.

Harvard Computation Laboratory (1955). *Tables of the Cumulative Binomial Probability Distribution*. Harvard University Press, Cambridge, Mass.

Haynes, J. D. (1948). An empirical investigation of sampling methods for an area. M. S. thesis, University of North Carolina.

Hendricks, W. A. (1944). The relative efficiencies of groups of farms as sampling units. *Jour. Amer. Stat. Assoc.*, **39**, 367–376.

Hendricks, W. A. (1949). Adjustment for bias by non-response in mailed surveys. *Agr. Econ. Res.*, **1**, 52–56.

Hendricks, W. A. (1956). *The Mathematical Theory of Sampling*. Scarecrow Press, New Brunswick, N.J.

Hess, I., Riedel, D. C., and Fitzpatrick, T. B. (1976). *Probability Sampling of Hospitals and Patients*. University of Michigan, Ann Arbor, Mich., second edition.

Hess, I., Sethi, V. K., and Balakrishnan, T. R. (1966). Stratification: A practical investigation. *Jour. Amer. Stat. Assoc.*, **61**, 74–90.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.*, **19**, 293–325.

Homeyer, P. G., and Black, C. A. (1946). Sampling replicated field experiments on oats for yield determinations. *Proc. Soil Sci. Soc. America*, **11**, 341–344.

Horvitz, D. G. (1952). Sampling and field procedures of the Pittsburgh morbidity survey. *Pub. Health Reports*, **67**, 1003–1012.

Horvitz, D. G., Greenberg, B. G., and Abernathy, J. R. (1975). Recent developments in randomized response designs. *A survey of statistical design and linear models*. J. N. Srivastava (ed.), American Elsevier Publishing Co., New York, 271–285.

Horvitz, D. G., Shah, B. V., and Simmons, W. R. (1967). The unrelated randomized response model. *Proc. Soc. Stat. Sect. Amer. Stat. Assoc.*, 65–72.

Horvitz, D. G., and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Jour. Amer. Stat. Assoc.*, **47**, 663–685.

Huddleston, H. F., Claypool, P. L., and Hocking, R. R. (1970). Optimum sample allocation to strata using convex programming. *App. Stat.* **19**, 273–278.

Hutchinson, M. C. (1971). A Monte Carlo comparison of some ratio estimators. *Biometrika*, **58**, 313–321.

Hyman, H. H. (1954). *Interviewing in Social Research*, University of Chicago Press, Chicago, Ill.

James, A. T., Wilkinson, G. N., and Venables, W. N. (1975). Interval estimates for a ratio of means. *Sankhya* (in press).

Jebe, E. H. (1952). Estimation for sub-sampling designs employing the county as a primary sampling unit. *Jour. Amer. Stat. Assoc.*, **47**, 49–70.

Jensen, A. (1926). Report on the representative method in statistics. *Bull. Int. Stat. Inst.*, **22**, 359–377.

Jessen, R. J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agr. Exp. Sta. Res. Bull.*, 304.

Jessen, R. J. (1955). Determining the fruit count on a tree by randomized branch sampling. *Biometrics*, **11**, 99–109.

Jessen, R. J., et al. (1947). On a population sample for Greece. *Jour. Amer. Stat. Assoc.*, **42**, 357–384.

Jessen, R. J., and Houseman, E. E. (1944). Statistical investigations of farm sample surveys taken in Iowa, Florida and California. *Iowa Agr. Exp. Sta. Res. Bull.*, 329.

Johnson, F. A. (1941). A statistical study of sampling methods for tree nursery inventories. M. S. thesis, Iowa State College.

Johnson, F. A. (1943). A statistical study of sampling methods for tree nursery inventories. *Jour. Forestry*, **41**, 674–689.

Jones, H. W. (1955). Investigating the properties of a sample mean by employing random subsample means. *Jour. Amer. Stat. Assoc.*, **51**, 54–83.

Kempthorne, O. (1969). Some remarks on inference in finite sampling. *New Developments in Survey Sampling*, N. L. Johnson and H. Smith, Jr. (eds.), John Wiley & Sons, New York, 671–695.

Kendall, M. G., and Smith, B. B. (1938). Randomness and random sampling numbers. *Jour. Roy. Stat. Soc.*, **101**, 147–166.

Keyfitz, N. (1957). Estimates of sampling variance where two units are selected from each stratum. *Jour. Amer. Stat. Assoc.*, **52**, 503–510.

Khan, S., and Tripathi, T. P. (1967). The use of multivariate auxiliary information in double-sampling. *J. Ind. Stat. Assoc.*, **5**, 42–48.

King, A. J., and McCarty, D. E. (1941). Application of sampling to agricultural statistics with emphasis on stratified samples. *Jour. Marketing*, April, 462–474.

Kiser, C. V., and Whelpton, P. K. (1953). Resume of the Indianapolis study of social and psychological factors affecting fertility. *Population Studies*, **7**, 95–110.

Kish, L. (1949). A procedure for objective respondent selection within the household. *Jour. Amer. Stat. Assoc.*, **44**, 380–387.

Kish, L. (1957). Confidence limits for clustered samples. *Amer. Soc. Rev.*, **22**, 154–165.

Kish, L. (1965). *Survey Sampling.* John Wiley & Sons, New York.

Kish, L., and Frankel, M. R. (1974). Inference from complex samples. *Jour. Roy. Stat. Soc.*, **B36**, 1–37.

Kish, L., and Hess, I. (1958). On noncoverage of sample dwellings. *Jour. Amer. Stat. Assoc.*, **53**, 509–524.

Kish, L., and Hess, I. (1959a). A "replacement" procedure for reducing the bias of nonresponse. *Amer. Statistician*, **13**, 4, 17–19.

Kish, L., and Hess, I. (1959b). On variances of ratios and their differences in multistage samples. *Jour. Amer. Stat. Assoc.*, **54**, 416–446.

Kish, L., and Lansing, J. B. (1954). Response errors in estimating the value of homes. *Jour. Amer. Stat. Assoc.*, **49**, 520–538.

Kish, L., Namboodiri, N. K., and Pillai, R. K. (1962). The ratio bias in surveys, *Jour. Amer. Stat. Assoc.*, **57**, 863–876.

Koch, G. (1973). An alternative approach to multivariate response error models for sample survey data with applications to estimators involving subclass means. *Jour. Amer. Stat. Ass.*, **68**, 906–913.

Kokan, A. R. (1963). Optimum allocation in multivariate surveys. *Jour. Roy. Stat. Soc*, **A126**, 557–565.

Koons, D. A. (1973). Quality control and measurement of nonsampling error in the Health Interview Survey. *Nat. Center for Health Stat.*, Series 2, 54.

Koop, J. C. (1960). On theoretical questions underlying the technique of replicated or interpenetrating samples. *Proc. Soc. Stat. Sect. Amer. Stat. Assoc.*, 196–205.

Koop, J. C. (1968). An exercise in ratio estimation. *Amer. Statistician*, **22**, 1, 29–30.

Kulldorff, G. (1963). Some problems of optimum allocation for sampling on two occasions. *Rev. Int. Stat. Inst.*, **31**, 24–57.

Lahiri, D. B. (1951). A method for sample selection providing unbiased ratio estimates. *Bull. Int. Stat. Inst.*, **33**, 2, 133–140.

Lieberman, G. J., and Owen, D. B. (1961). *Tables of the Hypergeometric Probability Distribution.* Stanford University Press, Stanford, Calif.

Lienau, C. C. (1941). Selection, training and performance of the National Health Survey field staff. *Amer. Jour. Hygiene*, **34**, 110–132.

Lund, R. E. (1968). Estimators in multiple frame surveys. *Proc. Soc. Sci. Sect. Amer. Stat. Assoc.*, 282–288.

McCarthy, P. J. (1966). Replication: An approach to the analysis of data from complex surveys. National Center for Health Statistics, Washington, D. C., Series, 2, 14.

McCarthy, P. J. (1969). Pseudo-replication: Half-samples. *Rev. Int. Stat. Inst.*, **37**, 239–264.

McVay, F. E. (1947). Sâmpling methods applied to estimating numbers of commercial orchards in a commercial peach area. *Jour. Amer. Stat. Assoc.*, **42**, 533–540.

Madow, L. H. (1946). Systematic sampling and its relation to other sampling designs. *Jour. Amer. Stat. Assoc.*, **41**, 207–214.

Madow, L. H. (1950). On the use of the county as a primary sampling unit for state estimates. *Jour. Amer. Stat. Assoc.*, **45**, 30–47.

Madow, W. G. (1948). On the limiting distributions of estimates based on samples from finite universes. *Ann. Math. Stat.*, **19**, 535–545.

Madow, W. G. (1949). On the theory of systematic sampling, II. *Ann Math. Stat.*, **20**, 333–354.

Madow, W. G. (1953). On the theory of systematic sampling, III. *Ann. Math. Stat.*, **24**, 101–106.

Madow, W. G., and Madow, L. H. (1944). On the theory of systematic sampling. *Ann. Math. Stat.*, **15**, 1–24.

Madow, W. G. (1965). On some aspects of response error measurement. *Proc. Soc. Stat. Soc. Amer. Stat. Assoc.*, 182–192.

Mahalanobis, P. C. (1944). On large-scale sample surveys. *Phil. Trans. Roy. Soc. London*, **B231**, 329–451.

Mahalanobis, P C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Jour. Roy. Stat. Soc.*, **109**, 325–370.

Matérn, B. (1947). Methods of estimating the accuracy of line and sample plot surveys. *Medd. fr. Statens Skogsforsknings Institut.*, **36**, 1–138.

Matérn, B. (1960). Spatial variation. *Medd. fr. Statens Skogsforsknings Institut.*, **49**, 5, 1–144.

Mickey, M. R. (1959). Some finite population unbiased ratio and regression estimators. *Jour. Amer. Stat. Assoc.*, **54**, 594–612.

Midzuno, H. (1951). On the sampling system with probability proportionate to sum of sizes *Ann. Inst. Stat. Math.*, **2**, 99–108.

Milne, A. (1959). The centric systematic area sample treated as a random sample. *Biometrics*, **15**, 270–297.

Moors, J. J. A. (1971). Optimization of the unrelated question randomized response model. *Jour. Amer. Stat. Assoc.*, **66**, 627–629.

Murthy, M. N. (1957). Ordered and unordered estimators in sampling without replacement. *Sankhya*, **18**, 379–390.

Murthy, M. N. (1967). *Sampling Theory and Methods.* Statistical Publishing Society, Calcutta, India.

Narain, R. D. (1951). On sampling without replacement with varying probabilities. *Jour. Ind. Soc. Agric. Stat.*, **3**, 169–174.

National Bureau of Standards (1950). *Tables of the Binomial Probability Distribution.* U.S. Government Printing Office, Washington, D.C.

Neter, J. (1972). How accountants save money by sampling. *Statistics, A Guide to the Unknown*, J. M. Tanur et al. (eds.), Holden-Day, Inc., San Francisco, 203–211.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Jour. Roy. Stat. Soc.*, **97**, 558–606.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *Jour. Amer. Stat. Assoc.*, **33**, 101–116.

Nordbotten, S. (1956). Allocation in stratified sampling by means of linear programming. *Skand. Akt. Tidskr.*, **39**, 1–6.

Nordin, J. A. (1944). Determining sample size. *Jour. Amer. Stat. Assoc.*, **39**, 497–506.

Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika*, **45**, 154–165.

Osborne, J. G. (1942). Sampling errors of systematic and random surveys of cover-type areas. *Jour. Amer. Stat. Assoc.*, **37**, 256–264.

Patterson, H. D. (1950). Sampling on successive occasions with partial replacement o[illegible] *Stat. Soc.*, **B12**, 241–255.

Patterson, H. D. (1954). The errors of lattice sampling. *Jour. Roy. Stat. Soc.*, **B16**, 140–149.

Paulson, E. (1942). A note on the estimation of some mean values for a bivariate distribution. *Ann. Math. Stat.*, **13**, 440–444.

Payne, S. L. (1951). *The Art of Asking Questions*. Princeton University Press, Princeton, N.J.

Plackett, R. L., and Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, **33**, 305–325.

Platek, R., and Singh, M. P. (1972). Some aspects of redesign of the Canadian Labor Force Survey. *Proc. Soc. Stat. Sect. Amer. Stat. Assoc.*, 397–402.

Politz, A. N., and Simmons, W. R. (1949, 1950). An attempt to get the "not at homes" into the sample without callbacks. *Jour. Amer. Stat. Assoc.*, **44**, 9–31, and **45**, 136–137.

Pritzker, L., and Hanson, R. (1962). Measurement errors in the 1960 Census of Population. *Proc. Soc. Stat. Sect. Amer. Stat. Assoc.*, 80–89.

Quenouille, M. H. (1949). Problems in plane sampling. *Ann. Math. Stat.*, **20**, 355–375.

Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, **43**, 353–360.

Raiffa, H., and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Harvard Business School, Cambridge, Mass.

Rand Corporation (1955). *A Million Random Digits*. Free Press, Glencoe, Ill.

Rao, C. R. (1971). Some aspects of statistical inference in problems of sampling from finite populations. *Foundations of Statistical Inference*. V. P. Godambe and D. A. Sprott, (eds.), Holt, Rinehart, and Winston, Toronto, Canada 177–202.

Rao, J. N. K. (1962). On the estimation of the relative efficiency of sampling procedures. *Ann. Inst. Stat. Math.*, **14**, 143–150.

Rao, J. N. K. (1965). On two simple schemes of unequal probability sampling without replacement. *Jour. Ind. Stat. Assoc.*, **3**, 173–180.

Rao, J. N. K. (1966). Alternative estimators in *pps* sampling for multiple characteristics. *Sankhya*, **A23**, 47–60.

Rao, J. N. K. (1968). Some small sample results in ratio and regression estimation. *Jour. Ind. Stat. Assoc.*, **6**, 160–168.

Rao, J. N. K. (1969). Ratio and regression estimators. *New Developments in Survey Sampling*, N. L. Johnson and H. Smith, Jr. (eds.), John Wiley & Sons, New York, 213–234.

Rao, J. N. K. (1973). On double sampling for stratification and analytical surveys. *Biometrika*, **60**, 125–133.

Rao, J. N. K. (1975a). On the foundations of survey sampling. In *A Survey of Statistical Design and Linear Models*, J. N. Srivastava (ed.), American Elsevier Publishing Co, New York, 489–505.

Rao, J. N. K. (1975b). Unbiased variance estimation for multistage designs. *Sankhya* (in press).

Rao, J. N. K., and Bayless, D. L. (1969). An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum. *Jour. Amer. Stat. Assoc.*, **64**, 540–559.

Rao, J. N. K., and Beegle, L. D. (1967). A Monte Carlo study of some ratio estimators. *Sankhya*, **B29**, 47–56.

Rao, J. N. K., and Graham, J. E. (1964). Rotation designs for sampling on repeated occasions. *Jour. Amer. Stat. Assoc.*, **59**, 492–509.

Rao, J. N. K., and Kuzik, R. A. (1974). Sampling errors in ratio estimation. *Indian Jour. Stat.* **36**, C, 43–58.

Rao, J. N. K., Hartley, H. O., and Cochran, W. G. (1962). A simple procedure of unequal probability sampling without replacement. *Jour. Roy. Stat. Soc.* **B24**, 482–491.

Rao, J. N. K., and Pereira, N. P. (1968). On double ratio estimators. *Sankhya*, **A30**, 83–90.

Rao, J. N. K., and Singh, M. P. (1973). On the choice of estimator in survey sampling. *Australian Jour. Stat.*, **15**, 95–104.

Rao, P. S. R. S., and Mudholkar, G. S. (1967). Generalized multivariate estimations for the mean of finite populations. *Jour. Amer. Stat. Assoc.*, **62**, 1008–1012.

Rao, P. S. R. S., and Rao, J. N. K. (1971). Small sample results for ratio estimators. *Biometrika*, **58,** 625–630.

Robson, D. S. (1952). Multiple sampling of attributes. *Jour. Amer. Stat. Assoc.*, **47,** 203–215.

Robson, D. S. (1957). Applications of multivariate polykays to the theory of unbiased ratio type estimation. *Jour. Amer. Stat. Assoc.*, **52,** 511–522.

Robson, D. S., and King, A. J. (1953). Double sampling and the Curtis impact survey. *Cornell Univ. Agr. Exp. Sta. Mem.*, 231.

Romig, H. G. (1952). *50–100 Binomial Tables*. John Wiley & Sons, New York.

Roy, J., and Chakravarti, I. M. (1960). Estimating the mean of a finite population. *Ann. Math. Stat.*, **31,** 392–398.

Royall, R. M. (1968). An old approach to finite population sampling theory. *Jour. Amer. Stat. Assoc.*, **63,** 1269–1279.

Royall, R. M. (1970a). On finite population sampling theory under certain linear regression models. *Biometrika*, **57,** 377–387.

Royall, R. M. (1970b). Finite population sampling—on labels in estimation., *Ann. Math. Stat.*, **41,** 1774–1779.

Royall, R. M. (1971). Linear regression models in finite population sampling theory. *Foundations of Statistical Inference*, V. P. Godambe, and D. A. Sprott (eds.), Holt, Rinehart, & Winston, Toronto, Canada, 259–279.

Royall, R. M., and Herson, J. (1973). Robust estimation in finite populations, I. *Jour. Amer. Stat. Assoc.*, **68,** 880–889.

Sagen, O. K., Dunham, R. E., and Simmons, W. R. (1959). Health statistics from record sources and household interviews compared. *Proc. Soc. Stat. Sect. Amer. Stat. Assoc.*, 6–15.

Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, **54,** 499–513.

Sandelius, M. (1951). Truncated inverse binomial sampling. *Skandinavisk Aktuarietidskrift*, **34,** 41–44.

Särndal, C. E. (1972). Sample survey theory vs. general statistical theory: Estimation of the population mean, *Rev. Int. Stat. Inst.*, **40,** 1–12.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics*, **2,** 110–114.

Scott, A. J., and Smith, T. M. F. (1974). Analysis of repeated surveys using time series methods. *Jour. Amer. Stat. Assoc.*, **69,** 674–678.

Sedransk, J. (1965). A double sampling scheme for analytical surveys. *Jour. Amer. Stat. Assoc.*, **60,** 985–1004.

Sedransk, J. (1967). Designing some multi-factor analytical studies. *Jour. Amer. Stat. Assoc.*, **62,** 1121–1139.

Sen, A. R. (1953). On the estimate of variance in sampling with varying probabilities. *Jour. Ind. Soc. Agric. Stat.*, **5,** 119–127.

Sen, A. R. (1972). Successive sampling with p ($p \geq 1$) auxiliary variables. *Ann. Math. Stat.*, **43,** 2031–2034.

Sen, A. R. (1973a). Theory and application of sampling on repeated occasions with several auxiliary variables. *Biometrics*, **29,** 383–385.

Sen, A. R. (1973b). Some theory of sampling on successive occasions. *Australian Jour. Stat.*, **15,** 105–110.

Seth, G. R., and Rao, J. N. K. (1964). On the comparison between simple random sampling with and without replacement. *Sankhya*, **A26,** 85–86.

Sethi, V. K. (1963). A note on optimum stratification for estimating the population means. *Australian Jour. Stat.*, **5,** 20–33.

Sethi, V. K. (1965). On optimum pairing of units. *Sankhya*, **B27,** 315–320.

Simmons, W. R. (1954). A plan to account for "not-at-homes" by combining weighting and callbacks. *Jour. of Marketing*, **11**, 42–53.

Singh, D., Jindal, K. K., and Garg, J. N., (1968). On modified systematic sampling. *Biometrika*, **55**, 541–546.

Sittig, J. (1951). The economic choice of sampling system in acceptance sampling. *Bull. Int. Stat. Inst.*, **33**, V, 51–84.

Slonim, M. J. (1960). *Sampling in a Nutshell.* Simon & Schuster, New York.

Smith, H. F. (1938). An empirical law describing heterogeneity in the yields of agricultural crops. *Jour. Agric. Sci.*, **28**, 1–23.

Smith, T. M. F. (1976). The foundations of survey sampling: A Review. *Jour. Roy. Stat. Soc.*, **A139**, 183–204.

Snedecor, G. W., and Cochran, W. G. (1967). *Statistical Methods.* Iowa State University Press, Ames, Iowa, sixth edition.

Srinath, K. P. (1971). Multiphase sampling in nonresponse problems. *Jour. Amer. Stat. Assoc.*, **16**, 583–586.

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Stat.*, **16**, 243–258.

Stephan, F. F. (1941). Stratification in representative sampling. *Jour. Marketing*, **6**, 38–46.

Stephan, F. F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoulli variate. *Ann. Math. Stat.*, **16**, 50–61.

Stephan, F., and McCarthy, P. J. (1958). *Sampling Opinions.* John Wiley and Sons, New York, p. 243.

Stuart, A. (1954). A simple presentation of optimum sampling results. *Jour. Roy. Stat. Soc.*, **B16**, 239–241.

Sukhatme, P. V. (1935). Contribution to the theory of the representative method. *Supp. Jour. Roy. Stat. Soc.*, **2**, 253–268.

Sukhatme, P. V. (1947). The problem of plot size in large-scale yield surveys. *Jour. Amer. Stat. Assoc.*, **42**, 297–310.

Sukhatme, P. V. (1954). *Sampling Theory of Surveys, With Applications.* Iowa State College Press, Ames, Iowa.

Sukhatme, P. V., and Seth, G. R. (1952). Non-sampling errors in surveys. *Jour. Ind. Soc. Agr. Stat.*, **4**, 5–41.

Sukhatme, P. V., and Sukhatme, B. V. (1970). *Sampling Theory of Surveys With Applications.* Food and Agriculture Organization, Rome, second edition.

Tepping, B. J., and Boland, K. L. (1972). Response variance in the Current Population Survey. U.S. Bureau of the Census Working Paper No. 36, U. S. Government Printing Office, Washington, D.C.

Tin, M. (1965). Comparison of some ratio estimators. *Jour. Amer. Stat. Assoc.*, **60**, 294–307.

Trueblood, R. M., and Cyert, R. M. (1957). *Sampling Techniques in Accounting.* Prentice-Hall, Englewood Cliffs, N.J.

Trussell, R. E., and Elinson, J. (1959). *Chronic Illness in a Large City.* Harvard University Press, Cambridge, Mass., pp. 339–370.

Tschuprow, A. A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, **2**, 461–493, 646–683.

Tukey, J. W. (1950). Some sampling simplified. *Jour. Amer. Stat. Assoc.*, **45**, 501–519.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples. *Ann. Math. Stat.*, **29**, 614.

U. N. Statistical Office (1950). *The preparation of sample survey reports.* Stat. Papers Series C, No. 1.

U. N. Statistical Office (1960). *Sample Surveys of Current Interest.* Eighth Report.

U. S. Bureau of the Census. (1968). *Evaluation and Research Program of the U.S. Census of Population and Housing, 1960: Effects of Interviews and Crew Leaders.* Series ER 60, No. 7, Washington, D.C.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Jour. Amer. Stat. Assoc.*, **60**, 63–69.

Warner, S. L. (1971). The linear randomized response model. *Jour. Amer. Stat. Assoc.*, **66**, 884–888.

Watson, D. J. (1937). The estimation of leaf areas. *Jour. Agr. Sci.*, **27**, 474.

West, Q. M. (1951). *The Results of Applying a Simple Random Sampling Process to Farm Management Data*. Agricultural Experiment Station, Cornell University.

Williams, W. H. (1963). The precision of some unbiased regression estimators. *Biometrika*, **17**, 267–274.

Wishart, J. (1952). Moment-coefficients of the k-statistics in samples from a finite population. *Biometrika*, **39**, 1–13.

Wold, H. O. A. (1954). *A Study in the Analysis of Stationary Time Series*. Almqvist and Wicksell, Stockholm, second edition.

Woodruff, R. S. (1959). The use of rotating samples in the Census Bureau's Monthly Surveys. *Proc. Soc. Stat. Sect. Amer. Stat. Assoc.*, 130–138.

Woodruff, R. S. (1971). A simple method for approximating the variance of a complicated estimate. *Jour. Amer. Stat. Assoc.*, **66**, 411–414.

Woolsey, T. D. (1956). Sampling methods for a small household survey. *Pub. Health Monographs*, No. 40.

Yates, F. (1948). Systematic sampling. *Phil. Trans. Roy. Soc. London*, **A241**, 345–377.

Yates, F. (1960). *Sampling Methods for Censuses and Surveys*. Charles Griffin and Co., London, third edition.

Yates, F., and Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Jour. Roy. Stat. Soc.*, **B15**, 253–261.

Zarkovic, S. S. (1960). On the efficiency of sampling with various probabilities and the selection of units with replacement. *Metrika*, **3**, 53–60.

Zukhovitsky, S. I., and Avdeyeva, L. I. (1966). *Linear and Convex Programming*. W. B. Saunders, Philadelphia.

ثبت المصطلحات

- أولاً : عربي - إنجليزي
- ثانياً : إنجليزي - عربي

أولاً : عربي - إنجليزي

أ

اتّساق	Consistensy
أثر التصميم	Design effect
احتمال متناسب مع الحجم	Probobility proportional to size
إحكام	Precision
اختبار إداري	Controlled selection
هادف	Purposive selection
أخطاء قياس	Errorsof measurements
ارتباط	Correlation
ما ضمن العنقود	Intracluster correlation
استبيان	Questionnaire
استخدامات مسوح العيّنة	Vises of sample surveys
استيخاء	Interpolation
إطار	Frame
إعادة مقابلة	Reinterview
أعداد عشوائية	Random numbers
أفضل تقدير خطي غير منحاز	Best linear unbiased estimate
التقدير النسبة المركب	Combined ratio estimate
المنفصل	Separate ratio estimate

التقدير النسبة متعدد المتغيرات	Multivariate ratio estimate
التواء	Skewness
الطريق الصعب (من غير المستجيبين)	Hard core (of nonrespendents)
المتوسط لكل وحدة	Mean per unit
المجتمع الهدف	Target population
المقدّر النسبة	Ratio estimator
انحراف معياري	Stadard deviation
انحياز	Bias

ب

بيانات إحصائية	Data

ت

تباين	Variance
استجابة	Response variance
كلي	Total response variance
تدقيق سجلات	Record checks
تصحيحات نهائية	End corrections
تصحيحي مجتمع منته	Finite population correction (FPC)
من أجل الاستمرار	Correction for continuity
تعداد (حصر شامل)	Consus
تغاير	Covariance
تغيير دوري	Periodic variation
تفرطح	Kurtosis
تقدير	Estimate
انحدار مركب	Combined regression estimate
منفصل	Separate regression estimate

Overestimate	تقدير بالزيادة
Eye estimate	بالعين المجرّدة
Variance estimate	تباين
Self-weighting estimate	ذاتي الترجيح
Unbiased estimate	غير منحاز
Proportion estimate	نسبة
Stratification	تقسيم إلى طبقات
Poststratification	تقدب إلى طبقات
Geographic stratification	جغرافي إلى طبقات
Travel cost	تكاليف السفر
Balaneed repeated replications	تكرارات معادة متوازنة
Binomial distribution	توزيع ثنائي (ذي الحدين)
Conditional distribution	شرطي
Normal distribution	طبيعي
Hypergymetric distribution	فوق الهندسي
Multinomial distribution	متعدد الحدود

ح

Optimum size of subsample	حجم أقل لعيّنة جزئية
Confidence limits	حد ثقة
Quadratic confidence limits	حدود ثقة تربيعية
Bounderies of strata	الطبقات
Response deviation	حيدان استجابة

خ

Standard error	خطأ معياري
Qualitative characteristics	خواص نوعية

د

دالة تباين	Variance function
تكرار	Frequency function
تكلفة	Cost function
خسارة	Loss function
مخاطرة	Risk function
درجات حرية	Degrees of freedom
دقة نسبية	Relative precision
دليل عدم اتساق	Indes of inconsistency
دمج تباينات	Pooling variances

ذ

ذاتية الارتباط	Autocorrelated

ز

زيارات متكررة	Call-backs

س

سؤال حساس	Sensitive question

ط

طبقات	Strata
منهارة	Collapsed strata
طبقة	Stratum
طريقة استجابة معمّاة	Randomized response method
سلاسل تايلور	Taylor series method
مدية الجيب	Jacknife method

ع

عامل تضخيم	Inflation factor
توسع	Expansion factor
عدّاد	Interviewer
عدم استجابة	Nonresponse
تغطية	Noncoverage
عنصر	Element
عيّنات جزئية متداخلة	Interpenetrating subsamples
عيّنة	Sample
شبكة مربعة	Squared grid sample
عكسية	Inverse sample
مساحية	Area sample
عطية مصفقة	Aligned systematic sample

غ

غير مستجيب	Nonresponded

ق

قاعدة $\sqrt{f}$ التجمعية	Comulative $\sqrt{f}$ rule
قياسات متكررة	Repaeated measurements

ك

كسر معاينة	Sampling fraction

ل

لا طبيعية	Non-normality

م

متوسط	Mean
مربعات خطأ	Mean square error
مجتمع	Population
جزئي	Subpopulation
فوقي	Superpopulation
محاصّة تناسبية	Proportional allocation
مثلى	Optimum allocation
نيمانية	Neyman allocation
مسح استطلاعي	Pilot survey
بالبريد	Mail survey
تحليلي	Analytical survey
عيّنة	Sample survey
مسوح وصفيّة	Descriptive surveys
مصوّر ارتباط	Corselogram
معامل اختلاف (تغيّر)	Coefficient of variation
ارتباط	Correlation coefficient
معاين	Sampler
معاينة	Sampling
احتمالية	Probability sampling
القبول	Acceptance sampling
بالحصة	Quata sampling
ثنائية الطور	Two phase sampling
دون إعادة	Sampling without replacement
شبكية	Lattice sampling
عشوائية	Random sampling

Simple random sampling	معاينة عشوائية بسيطة
Stratified random sampling	طبقية
Three stage sampling	على ثلاث مراحل
Two stage sampling	على مرحلتين
Cluster samping	عنقودية
Multistage sampling	متعددة المراحل
Repeated sampling	متكررة
Double sampling	مضاعفة
Sampling with replacement	مع الإعادة
Systematic sampling	نمطية
Unaligned systematic sampling	غير مصففة
Rare items	مفردات نادرة
Item	مفردة
Estimator	مقدّر
Regression estimator	انحدار
Linear regression estimator	خطي
Product estimator	جدائي
Measure of homogeneity	مقياس تجانس
Domains of study	ميادين دراسة

ن

proportion	نسبة
Percentage	مئوية
Simple expansion	نشر بسيط
Central limit theorem	نظرية النهاية المركزية
Unbiased moded	نموذج لا انحياز

و

وحدة	Unit
فرعية	Subunit
معاينة	Sampling unit
أولية	Primary sampling unit

ثانيًا: إنجليزي - عربي

A

English	عربي
Acceptance sampling	معاينة القبول
Aligned systematic sample	عيّنة نمطية مصففة
Analytical survey	مسح تحليلي
Area sample	عيّنة مساحية
Autocorrelated	ذاتية الارتباط

B

English	عربي
Balanced repeated replications	تكرارات معادة متوازنة
Best linear unbiased estimate	أفضل تقدير خطي غير منحاز
Bias	انحياز
Binomial distribution	توزيع ثنائي (ذي الحدين)
Bounderies of strata	حدود الطبقات

C

English	عربي
Call-backs	زيارات متكررة
Census	تعداد (حصر شامل)
Central limit theorem	نظرية النهاية المركزية
Cluster sampling	معاينة عنقودية
Coefficient of variation	معامل اختلاف (تغير)
Collapsed strata	طبقات منهارة

English	Arabic
Combined ratio estimate	التقدير النسبة المركب
regression estimate	تقدير انحدار مركب
Conditional distribution	توزيع شرطي
Confidence limits	حد الثقة
Consistency	اتّساق
Controlled selection	اختيار إرادي
Correction for continuity	تصحيح من أجل الاستمرار
Correlation	ارتباط
coefficient	معامل ارتباط
Correlogram	مصوّر ارتباط
Cost function	دالة متكلفة
Covariance	تغاير
Cumulative $\sqrt{f}$ rule	قاعدة $\sqrt{f}$ التجمعية

D

English	Arabic
Data	بيانات إحصائية
Degrees of freedom	درجات حرية
Descriptive surveys	مسوح وصفية
Design effect (Deff.)	أثر التصميم
Domains of study	ميادين دراسة
Double sampling	معاينة مضاعفة

E

English	Arabic
Element	عنصر
End corrections	تصحيحات نهائية
Errors of measurement	أخطاء قياس
Estimate	تقدير
Estimator	مقدّر

Expansion factor	عامل توسّع
Eye estimate	تقدير بالعين المجرّدة

F

Finite population correction (FPC)	تصحيح مجتمع منته (ت م م)
Frame	إطار
Frequency Function	دالة تكرار

G

Geographic stratification	تقسيم جغرافي إلى طبقات

H

Hard core (of nonrespondents)	الفريق الصعب (من غير المستجيبين)
Hypergeometric distribution	توزيع فوق الهندسي

I

Index of inconsistency	دليل عدم اتّساق
Inflation factor	عامل تضخيم
Interpenetrating subsamples	عيّنات جزئية متداخلة
Interpolation	استيفاء
Intracluster correlation	ارتباط ما ضمن العنقود
Inverse sample	عيّنة عكسية
Item	مفردة

J

Jacknife method	طريقة مدية الجيب

K

Kurtosis	تفرطح

L

Lattice sampling	معاينة شبكية
Linear regression estimator	مقدر انحدار خطي

Loss function	دالة خسارة

M

Mail survey	مسح بالبريد
Mean	متوسط
per unit	المتوسط لكل وحدة
square error	متوسط مربعات خطأ
Measure of homogeneity	مقياس تجانس
Model - unbiased	نموذج ـ لا انحياز
Multinomial distribution	توزيع متعدد الحدود
Multistage sampling	معاينة متعددة المراحل
Multivariate ratio estimate	التقدير النسبة متعدد المتغيرات

N

Neyman allocation	محاصّة نيمانية
Noncoverage	عدم تغطية
Non-normality	لا طبيعية
Nonresponce	عدم استجابة
Nonrespondent	غير مستجيب
Normal distribution	توزيع طبيعي

O

Optimum allocation	محاصّة مثلى
size of subsample	حجم أمثل لعيّنة جزئية
Over estimate	تقدير بالزيادة

P

Percentage	نسبة مئوية
Periodic variation	تغير دوري
Pilot survey	مسح استطلاعي

Pooling variances	دمج تباينات
Population	مجتمع
Poststratification	تقسيم بَعْدي إلى طبقات
Precision	إحكام
Pretest	اختبار مسبق
Primary sampling unit	وحدة معاينة أولية
probability proportional to size	احتمال متناسب مع الحجم
sampling	معاينة احتمالية
Product estimator	مقدّر جُدائي
Proportion	نسبة
Proportional allocation	محاصّة تناسبية
Proportion estimate	تقدير نسبة
Porposive selection	اختيار هادف

Q

Quadratic confidence limits	حدود ثقة تربيعية
Qualitative charecteristics	خواص نوعية
Questionnaire	استبيان
Quota sampling	معاينة بالحصّة

R

Randomized response method	طريقة استجابة معشّاة
Random numbers	أعداد عشوائية
sampling	معاينة عشوائية
Rare items	مفردات نادرة
Ratio estimator	المقدّر النسبة
Record checks	تدقيق سجلات
Regression estimator	مقدّر انحدار

Reinterview	إعادة مقابلة
Relative precision	دقة نسبية
Repeated measurements	قياسات متكررة
sampling	معاينة متكررة
Response deviation	حيدان استجابة
variance	تباين استجابة
Risk function	دالة مخاطرة

S

Sample	عيّنة
Sampler	معاين
Sample survey	مسح عيّنة
Sampling	معاينة
fraction	كسر معاينة
unit	وحدة معاينة
without replacement	معاينة دون إعادة
'with replacement	معاينة مع إعادة
Self-weighting estimate	تقدير ذاتي الترجيح
Sensitive question	سؤال حساس
Separate ratio estimate	التقدير النسبة المنفصل
regression estimate	تقدير انحدار منفصل
Simple expansion	نشر بسيط
random sanpling	معاينة عشوائية بسيطة
Skewness	التواء
Square grid sample	عيّنة شبكة مربعة
Standard deviation	انحراف معياري
error	خطأ معياري

Strata	طبقات
Stratification	تقسيم إلى طبقات
Stratified random samlling	معاينة عشوائية طبقية
Stratum	طبقة
Subpopulation	مجتمع جزئي
Subunit	وحدة فرعية (جزئية)
Superpopulation	مجتمع فوقي
Systematic sampling	معاينة نمطية

T

Target population	المجتمع الهدف
Taylor series method	طريقة سلاسل تايلور
Three stage sampling	معاينة على ثلاث مراحل
Total response variance	تباين استجابة كلي
Travel costs	تكاليف السفر
Two-phase sampling	معاينة ثنائية الطور (مضاعفة)
Two-stage sampling	معاينة على مرحلتين

U

Unaligned systematic sampling	معاينة نمطية غير مصففة
Unbiased estimate	تقدير غير منحاز
Unit	وحدة
Uses of sample surveys	استخدامات مسوح العيّنة

V

Variance	تباين
estimate	تقدير تباين
function	دالة تباين

كشّاف الموضوعات

ا

مـ

و

ردمك: ٥-٢٥٧-٠٥-٩٩٦٠
ISBN:9960-05-257-5